Estimating Exposome Score for Schizophrenia Using Predictive Modeling Approach in Two Independent Samples: The Results From the EUGEI Study

Lotta-Katrin Pries[1], Agustin Lage-Castellanos[2,3], Philippe Delespaul[1], Gunter Kenis[1], Jurjen J. Luykx[4–6], Bochao D. Lin[5], Alexander L. Richards[7], Berna Akdede[8], Tolga Binbay[8], Vesile Altinyazar[9], Berna Yalinçetin[10], Güvem Gümüş-Akay[11], Burçin Cihan[12], Haldun Soygür[13], Halis Ulaş[14], Eylem Şahin Cankurtaran[15], Semra Ulusoy Kaymak[16], Marina M. Mihaljevic[17,18], Sanja Andric Petrovic[18], Tijana Mirjanic[19], Miguel Bernardo[20–22], Bibiana Cabrera[20,22], Julio Bobes[22–25], Pilar A. Saiz[22–25], María Paz García-Portilla[22–25], Julio Sanjuan[22,26], Eduardo J. Aguilar[22,26], José Luis Santos[22,27], Estela Jiménez-López[22,28], Manuel Arrojo[29], Angel Carracedo[30], Gonzalo López[22,31], Javier González Peñas[22,31], Mara Parellada[22,31], Nadja P. Maric[17,18], Cem Atbaşoğlu[32], Alp Ucok[33], Köksal Alptekin[8], Meram Can Saka[32]; Genetic Risk and Outcome of Psychosis (GROUP) investigators[†], Celso Arango[22,31], Michael O'Donovan[7], Bart P.F. Rutten[1,36], Jim van Os[1,4,34,36], and Sinan Guloksuz[*,1,35,36]

1 Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Center, Maastricht, The Netherlands;
2 Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands;
3 Department of NeuroInformatics, Cuban Center for Neuroscience, Havana, Cuba;
4 Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;
5 Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;
6 GGNet Mental Health, Apeldoorn, The Netherlands;
7 MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK;
8 Department of Psychiatry, Dokuz Eylul University School of Medicine, Izmir, Turkey;
9 Department of Psychiatry, Faculty of Medicine, Adnan Menderes University, Aydin, Turkey;
10 Department of Neuroscience, Health Sciences Institute, Dokuz Eylul University, Izmir, Turkey;
11 Ankara University Brain Research Center, Ankara, Turkey;
12 Department of Psychology, Middle East Technical University, Ankara, Turkey;
13 Turkish Federation of Schizophrenia Associations, Ankara, Turkey;
14 Dokuz Eylül University, Medical School, Psychiatry Department (Discharged from by statutory decree No:701 at 8th July of 2018 because of signing "Peace Petition")
15 Güven Çayyolu Healthcare Campus, Ankara, Turkey;
16 Atatürk Research and Training Hospital Psychiatry Clinic, Ankara, Turkey;
17 Faculty of Medicine, University of Belgrade, Belgrade, Serbia;
18 Clinic for Psychiatry CCS, Belgrade, Serbia;
19 Special Hospital for Psychiatric Disorders Kovin, Kovin, Serbia;
20 Barcelona Clinic Schizophrenia Unit, Neuroscience Institute, Hospital Clinic of Barcelona, University of Barcelona, Barcelona, Spain;
21 Institutd'Investigacions Biomèdiques August Pi I Sunyer, Barcelona, Spain;
22 Biomedical Research Networking Centre in Mental Health (CIBERSAM), Spain;
23 Department of Psychiatry, School of Medicine, University of Oviedo, Oviedo, Spain;
24 Instituto de Investigación Sanitaria del Principado de Asturias, Oviedo, Spain;
25 Mental Health Services of Principado de Asturias, Oviedo, Spain;
26 Department of Psychiatry, Hospital Clínico Universitario de Valencia, School of Medicine, Universidad de Valencia, Valencia, Spain;
27 Department of Psychiatry, Hospital Virgen de la Luz, Cuenca, Spain;
28 Universidad de Castilla-La Mancha, Health and Social Research Center, Cuenca, Spain;
29 Department of Psychiatry, Instituto de Investigación Sanitaria, Complejo Hospitalario Universitario de Santiago de Compostela, Santiago de Compostela, Spain;
30 Fundación Publica Galega de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, Spain;

31 Department of Child and Adolescent Psychiatry, Hospital General Universitario Gregorio Marañón, IiSGM, School of Medicine, Universidad Complutense, Madrid, Spain;
32 Department of Psychiatry, School of Medicine, Ankara University, Ankara, Turkey;
33 Department of Psychiatry, Faculty of Medicine, Istanbul University, Istanbul, Turkey;
34 Department of Psychosis Studies, King's College London, Institute of Psychiatry, London, UK;
35 Department of Psychiatry, Yale School of Medicine, New Haven, CT
36 These authors contributed equally to the article.

†The Genetic Risk and Outcome of Psychosis (GROUP) investigators in EUGEI are listed in the Appendix.

*To whom correspondence should be addressed; Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Center, P.O. Box 616 6200 MD Maastricht, The Netherlands; tel: 31-433-88-4071, fax: 31- 433-88-4122, e-mail: sinan.guloksuz@maastrichtuniversity.nl

Exposures constitute a dense network of the environment: exposome. Here, we argue for embracing the exposome paradigm to investigate the sum of nongenetic "risk" and show how predictive modeling approaches can be used to construct an exposome score (ES; an aggregated score of exposures) for schizophrenia. The training dataset consisted of patients with schizophrenia and controls, whereas the independent validation dataset consisted of patients, their unaffected siblings, and controls. Binary exposures were cannabis use, hearing impairment, winter birth, bullying, and emotional, physical, and sexual abuse along with physical and emotional neglect. We applied logistic regression (LR), Gaussian Naive Bayes (GNB), the least absolute shrinkage and selection operator (LASSO), and Ridge penalized classification models to the training dataset. ESs, the sum of weighted exposures based on coefficients from each model, were calculated in the validation dataset. In addition, we estimated ES based on meta-analyses and a simple sum score of exposures. Accuracy, sensitivity, specificity, area under the receiver operating characteristic, and Nagelkerke's R^2 were compared. The $ES_{Meta\text{-}analyses}$ performed the worst, whereas the sum score and the ES_{GNB} were worse than the ES_{LR} that performed similar to the ES_{LASSO} and ES_{RIDGE}. The ES_{LR} distinguished patients from controls (odds ratio [OR] = 1.94, $P < .001$), patients from siblings (OR = 1.58, $P < .001$), and siblings from controls (OR = 1.21, $P = .001$). An increase in ES_{LR} was associated with a gradient increase of schizophrenia risk. In reference to the remaining fractions, the ES_{LR} at top 30%, 20%, and 10% of the control distribution yielded ORs of 3.72, 3.74, and 4.77, respectively. Our findings demonstrate that predictive modeling approaches can be harnessed to evaluate the exposome.

Key words: schizophrenia/psychosis/predictive modeling/ machine learning/risk score/environment/childhood trauma/cannabis/winter birth/hearing impairment

Introduction

Several environmental exposures have been associated with psychosis spectrum disorder.[1,2] Knowledge on this association has thus far been deduced from hypothesis driven selective one-exposure to one-outcome studies, akin to the candidate-gene approach.[3] However, each exposure constitutes a fraction of a dense network of exposures: the exposome.[4] Here, we argue for embracing the exposome paradigm to investigate the sum of the nongenetic "risk" and show how a predictive modelling approach can be used to construct an exposome score (ES) for schizophrenia, a single metric of aggregated environmental load similar to polygenic risk score.[5]

Approach

Guided by the predictive modeling methods for constructing cumulative environmental exposure scores,[6,7] we used 2 independent datasets to, first, build a predictive model in the training dataset (the Work-package 6 of the European Network of National Networks studying Gene-Environment Interactions in Schizophrenia [EUGEI][2]) and, second, construct and test the ES in the validation dataset (the Genetic Risk and Outcome of Psychosis [GROUP] study[8]). We examined the following widely evaluated environmental factors that we also recently investigated individually within the context of gene-environment interaction[9]: hearing impairment, winter birth, cannabis use, and childhood adversities (bullying, emotional, physical, and sexual abuse along with emotional and physical neglect).[10] Our analysis was limited to the environmental exposures that were reliably measured and equally available in both datasets. These environmental factors were defined according to previous studies.[9] The detailed description of each environmental exposure is provided in the

supplementary file. We used 4 prediction models to determine to what degree cumulative environmental exposure contributes to the liability for schizophrenia in a case-control design. Logistic regression (LR), Gaussian Naive Bayes (GNB), and penalized logistic regression (least absolute shrinkage and selection operator [LASSO] and Ridge) were applied to data with complete information on environmental exposures. The description of the models and the distribution of exposures are provided in the supplementary file. For each model, the dependent variable was the binary case-control status, whereas binary environmental exposures were features (independent variables). First, we estimated coefficients of binary exposures in the training dataset including 1241 healthy controls and 747 patients with a diagnosis of schizophrenia spectrum disorders. Second, we calculated the weighted sum of the exposures according to each predictive model in an independent validation dataset with 323 healthy controls, 463 patients with a diagnosis of schizophrenia spectrum disorders, and 542 unaffected siblings of the patients. To compare the performance of ES from each model, we also generated an environmental sum score by simply adding each binary exposure per individual as 0 = absent and 1 = present (the sum score is ranging from 0 to 9) and a cumulative environmental score weighted by the metanalytical estimates for each exposure,[11–14] conforming to a previous study.[15] Finally, we tested the performance of ESs derived from each model by applying logistic regression in a case-control design in the independent validation dataset by evaluating the area under the receiver operating characteristic (ROC), accuracy (ACC), sensitivity, specificity, and Nagelkerke's pseudo R_2. In this regard, we prioritized models with better sensitivity than specificity as our main concern was to avoid misclassifying individuals diagnosed with schizophrenia.

Prediction in the Training Dataset

The coefficients of individual models (see figure 1a and supplementary table S2) indicate that cannabis use (coefficients ranging from 1.31 to 1.53), hearing impairment (coefficients: 1.10–1.19), and bullying (coefficients: 1.30–1.57) received the highest weights in the training dataset. The lowest weight was attributed to winter birth with coefficients between 0.01 and 0.06. In comparison with the GNB model, which assumes independence between predictors, the LR, Ridge, and LASSO models yielded lower weights for emotional abuse, sexual abuse, emotional neglect, physical neglect, and bullying. Further, although physical abuse was a strong positive predictor in the GNB model, its predictive value was lost and even yielded a negative weight when using predictive model approaches that account for dependence between the predictors. This is in line with evidence that exposures are weakly to moderately correlated with each other.[3,16–18] Consequently, coefficients are overestimated when independence is assumed.

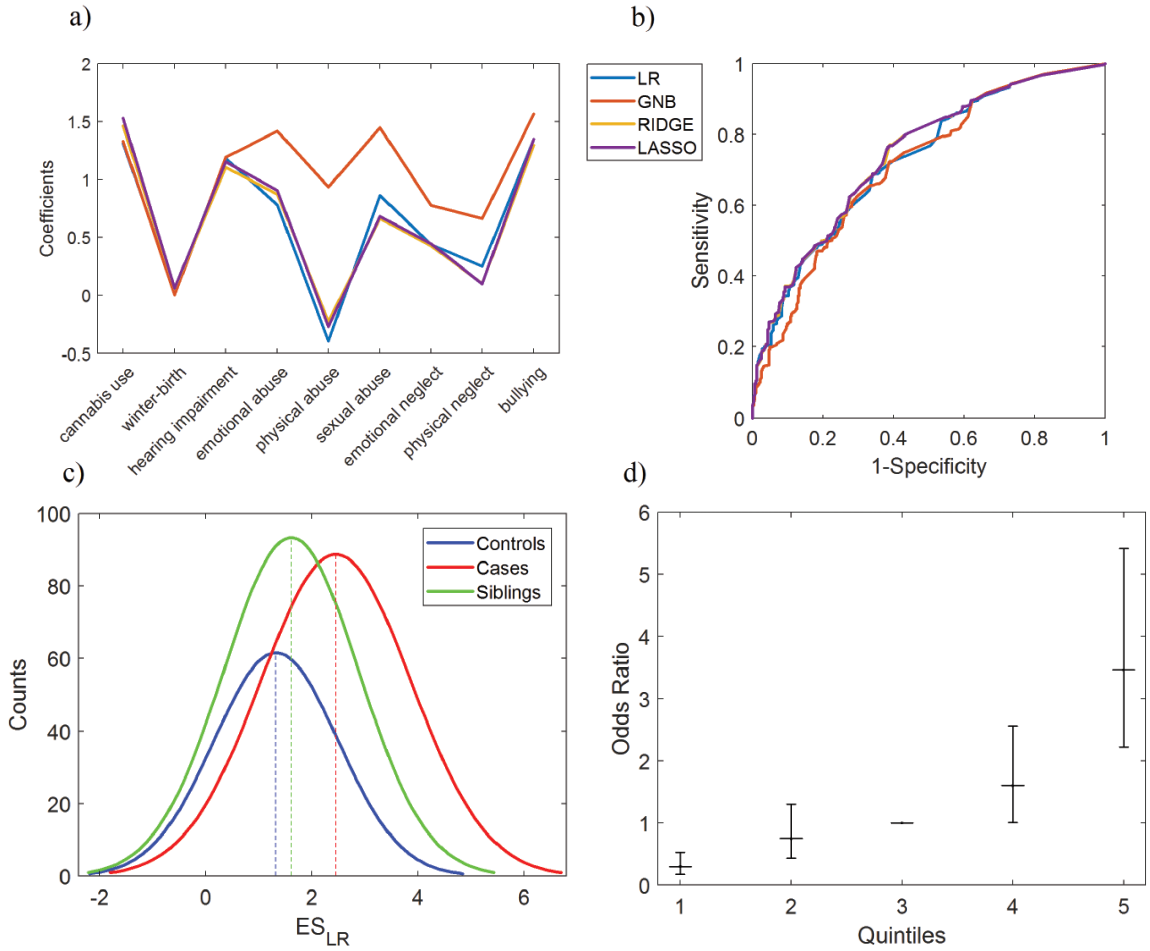

Fig. 1. (a) Coefficients profile for each exposure derived from different classification methods in the training dataset, GNB: Gaussian Naive Bayes, LR: logistic regression. (b) The area under the receiver operating characteristic for the different exposome scores in the validation dataset. (c) The histogram of the ES_{LR} (exposome score based on logistic regression) for patients, siblings, and controls in the validation dataset. For visualization, a Gaussian distribution was fit to histogram counts by adjusting mean and standard deviations. (d) The risk strata plot of the ES_{LR} on case-control status: The ES_{LR} was divided into 5 quintiles (X-axis) of the control distribution and logistic regression was applied to case-control status as the dependent variable. The third quintile includes the median and was used as reference. The Y-axis represents odds ratios and the error bars show confidence intervals.

Constructing and Testing the Performance of Exposome Score in an Independent Dataset

The ROC was used to estimate the performance of the calculated ESs in predicting the case-control status in the validation dataset (figure 1b and supplementary table S3). The ES based on meta-analytical estimates (the $ES_{Meta\text{-}analyses}$), ES_{GNB}, and the environmental sum score yielded the lowest ROC, 0.69, 0.71, and 0.71, respectively, whereas all other ESs (ES_{LR}, ES_{RIDGE}, and ES_{LASSO}) had ROC ranging from 0.73 to 0.74. With a chance level of 0.5 (as patients and controls were in balance in the training sample; see supplementary file), all ESs indicated an ACC above chance level (ACC: 0.62–0.68) with specificity between 0.42 and 0.72 and sensitivity between 0.56 and 0.86. Compared to the ES_{LR}, ES_{RIDGE}, and ES_{LASSO}, the ESs derived from the models assuming independence between exposures (ES_{GNB}, environmental sum score, and $ES_{Meta\text{-}analyses}$) performed worse on sensitivity and had more false negatives as they incorrectly classified patients as healthy. Given that our priority was reducing false negatives rather than reducing false positives and that the ES_{LR}, ES_{RIDGE}, and ES_{LASSO} performed similarly well (figure 1b and supplementary table S3), we reported further analyses with the ES_{LR}, which was constructed on the basis of a widely available and commonly used statistical model, logistic regression.

To examine whether the ES_{LR} reflects schizophrenia liability in the validation dataset, we evaluated the ES_{LR} in patients, siblings, and controls (see figure 1c for an illustration and supplementary table S4 for the other models). The ES_{LR} discriminated patients from controls (odds ratio [OR] = 1.94; 95% confidence interval [CI] = 1.71–2.20; $P < .001$, Nagelkerke's pseudo $R_2 = 0.21$), also after adjusting for age and sex (OR = 1.87; 95% CI = 1.64–2.14; $P < .001$) in the validation dataset. Similarly, logistic regression

analysis showed higher ES_{LR} in patients compared to siblings (OR = 1.58; 95% CI = 1.43–1.74; $P < .001$; adjusted for age and sex: OR = 1.55; 95% CI = 1.40–1.72; $P < .001$) and in siblings compared to controls (OR = 1.21; 95% CI = 1.08–1.36; $P = .001$; adjusted for age and sex: OR = 1.23; 95% CI = 1.09–1.38; $P < .001$).

To visually represent the risk stratification properties of the ES_{LR}, we categorized the ES_{LR} using the quintiles of the control distribution and measured the case-control ORs using the middle quintile (median ES_{LR}) as the reference. With an increase of the ES_{LR}, we noticed a gradient increase in the risk for schizophrenia. In comparison with the median, the fifth quintile had a higher OR (OR = 3.47; 95% CI = 2.22–5.41; $P < .001$ and age- and sex-adjusted OR = 3.78; 95% CI = 2.34–6.09; $P < .001$) and the first quintile had a lower OR (OR = 0.30; 95% CI = 0.17–0.53; $P < .001$ and age- and sex-adjusted OR = 0.34; 95% CI = 0.19–0.62; $P < .001$; figure 1c). We then dichotomized the ES_{LR} with cut off points at 70%, 80%, and 90% of the control distribution. Comparing the top and the bottom part translated to ORs of 3.81, 3.96, and 5.11 (age- and sex-adjusted ORs of 3.72, 3.74, and 4.77) for 70%, 80%, and 90% of the distribution, respectively (supplementary table S5).

Discussion

For the first time, we applied a predictive modelling approach to construct the ES for schizophrenia by leveraging 2 large independent datasets (training and validation data) with similar assessment protocols for environmental exposures. Our findings suggest that predictive modeling can be used to estimate environmental loading of a range of exposures. We found that the ES_{LR}, ES_{RIDGE}, and ES_{LASSO} performed similarly well, whereas the ESs derived from the models assuming independence performed worse. Of the ES_{GNB}, $ES_{Meta\text{-}analyses}$, and the simple summation of exposures, the $ES_{Meta\text{-}analyses}$, relying on the external sources for extracting estimates for environmental exposures, showed the worst performance. The low performance of the ES driven by meta-analyses might be related to the fact that meta-analytical estimates are derived from different studies that use different assessments, different definitions, and different cutoff points for exposures in different study populations,[3] which might not be completely compatible with the dataset at hand. The availability of similar training and validation datasets plays a major role in prediction power—for instance, the predictive performance of polygenic scores for schizophrenia is considerably lower in non-Caucasian ancestry samples.[19] Therefore, a similar situation exists in estimating genetic liability, which, however, has the advantage of using more concrete, uniformly measured genetic variation for prediction in comparison to environmental assessment. Generating a uniform "environmental risk score" is even more challenging. For instance, cannabis use could be scored positive if participants smoke daily, or at least weekly, or at least monthly for lifetime use or exposure during adolescence, whereas childhood adversities could similarly be measured by various methods. Therefore, as weights are determined by how strict or lenient the cutoff points are, it is likely that the inconsistency between sampling and measurement strategies would introduce bias. Further, when individual coefficients from meta-analyses are used for a weighted environmental score, correlations between exposures are ignored, and weights may be overestimated.[3] In line with this, we also show that GNB, which assumes independence between predictors, produces higher weights for exposures than the other data-driven models.

Similar to current results, previous studies show that more contemporary algorithms do not necessarily translate into superior performance over logistic regression for clinical prediction modeling.[20,21] However, it should be noted that our analysis did not involve a complex data structure with many predictors. Penalized classification models might have led to performance improvement if more complex structures had to be considered (eg, increasing the number of predictors and adding pairwise interactions). Researchers likewise need to be cautious about overfitting models and be aware that, if environmental exposures are correlated, the initial simple model with a few predictors will show the highest portion of improvement. However, each sequentially added predictor would result in less and less improvement in model performance.[21]

The ESs assuming independence between predictors (sum score, $ES_{Meta\text{-}analyses}$, and ES_{GNB}) had lower sensitivity than the rest. The $ES_{Meta\text{-}analyses}$ indicated the lowest sensitivity (56%). The sensitivity of an environmental score derived from meta-analytical estimates in a previous study was even lower, only around 7%–9%.[15] In other words, predictive models that do not assume independence between exposures may more accurately classify patients as positive by decreasing false negatives. The ESs from the models assuming independence, however, had higher specificity and were better in decreasing false positives. As our main concern was to avoid misclassifying individuals diagnosed with schizophrenia, we chose sensitivity over specificity. Further, if more environmental exposures were to be included in the ES, thus introducing more correlation, the models not assuming independence between predictors (ES_{LR}, ES_{RIDGE}, and ES_{LASSO}) would perform increasingly better than the models assuming independence

(sum score, $ES_{Meta\text{-}analyses}$, and ES_{GNB}).
The ES_{LR}, generated using an easily accessible method (logistic regression), achieved similar performance results compared with the ES_{RIDGE} and ES_{LASSO}. We used the ES_{LR} to further explore the characteristics of the ES in the follow- up analyses. In general, patients had higher ES_{LR} than both controls and siblings, whereas siblings had higher ES_{LR} than controls. The ES_{LR} explained more variance (Nagelkerke $R^2 = 0.21$) than the $ES_{Meta\text{-}analyses}$ (Nagelkerke $R^2 = 0.13$). In accordance with our previous findings showing an additive effect for environmental factors,[22,23] our results indicate that the ES_{LR} shows a dose-response effect: the odds of schizophrenia increase as a function of the ES_{LR}. Eventually, an individual with ES_{LR} in the top 10% of the control distribution was around 5 times more likely to have schizophrenia compared to an individual below that cutoff.

Limitations of Exposome Score
Our analysis was limited to the environmental exposures that were reliably measurable and equally available in both datasets. The ES can be extended to include other environmental exposures (eg, obstetric and pregnancy complications and urban environment). We included winter birth as an exposure in the current analyses as previous studies suggest an association between winter birth and psychosis.[14] However, summer birth was also previously associated with deficit schizophrenia and might therefore be evaluated as an exposure as well.[24,25] Considering evidence showing that common environmental factors (eg, childhood adversity) are not specific to the psychosis phenotype but instead are more generally related to psychopathology,[26,27] the ES would likely (to a degree) be associated with other mental disorders in mixed samples. Therefore, a low discriminant capacity for the ES should be anticipated. Given the nature of observational studies, causality claims should be avoided. Finally, it should be noted that although aggregating exposures leads to an increase in the predictive power and may be particularly beneficial in exploring shared mechanisms, the inherent heterogeneity of a single score may lead to information loss and biological imprecision. Considering the reasons described earlier, we have avoided using the term "risk" and opted for a neutral alternative: ES.

Conclusion
Our findings demonstrate that predictive modelling approaches can be harnessed to evaluate the exposome. In the future, we aim to explore models by including more exposures as well as interaction terms and test the predictive power of the ES in epidemiologically representative general population cohorts.

Funding
The EUGEI project was supported by the grant agreement HEALTH-F2-2010-241909 from the European Community's Seventh Framework Programme. The authors are grateful to the patients and their families for participating in the project. They also thank all research personnel involved in the GROUP project, in particular J. van Baaren, E. Veermans, G. Driessen, T. Driesen, E. van't Hag and J. de Nijs. Bart PF Rutten was funded by a VIDI award number 91718336 from the Netherlands Scientific Organisation.

Appendix

GROUP-EUGEI investigators are: Behrooz Z. Alizadeh[1], Therese van Amelsvoort[2], Richard Bruggeman[1], Wiepke Cahn[3,4], Lieuwe de Haan[5], Jurjen J. Luykx[3,6,7], Ruud van Winkel[2,8], Bart P.F. Rutten[2], Jim van Os[2,3,9]

[1]University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Rob Giel Research center, Groningen, The Netherlands;
[2]Maastricht University Medical Center, Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht, The Netherlands;
[3]Department of Psychiatry, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands;
[4]Altrecht, General Menthal Health Care, Utrecht, The Netherlands;
[5]Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands;
[6]Department of Translational Neuroscience, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands;
[7]GGNet Mental Health, Apeldoorn, The Netherlands;
[8]KU Leuven, Department of Neuroscience, Research Group Psychiatry, Leuven, Belgium King's;
[9]College London, King's Health Partners, Department of Psychosis Studies, Institute of Psychiatry, London, United Kingdom

References

1. van Os J, Kenis G, Rutten BP. The environment and schizophrenia. *Nature.* 2010;468(7321):203–212.
2. EUGEI. Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations. *Schizophr Bull.* 2014;40:729–736.
3. Guloksuz S, Rutten BPF, Pries LK, et al.; European Network of National Schizophrenia Networks Studying Gene-Environment Interactions Work Package 6 (EU-GEI WP6) Group. The complexities of evaluating the exposome in psychiatry: a data-driven illustration of challenges and some propositions for amendments. *Schizophr Bull.* 2018;44(6):1175–1179.
4. Guloksuz S, van Os J, Rutten BPF. The exposome paradigm and the complexities of environmental research in psychiatry. *JAMA Psychiatry.* 2018;75(10):985–986.
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophreniaassociated genetic loci. *Nature.* 2014;511:421–427.
6. Oulhote Y, Bind M-A, Coull B, Patel CJ, Grandjean P. 2017. Combining ensemble learning techniques and G-Computation to investigate chemical mixtures in environmental epidemiology studies. *Biorxiv.* 2017;147413. First published on June 30, 2017, doi: 10.1101/147413
7. Park SK, Zhao Z, Mukherjee B. Construction of environmental risk score beyond standard linear models using machine learning methods: application to metal mixtures, oxidative stress and cardiovascular disease in NHANES. *Environ Health.* 2017;16(1):102.
8. Korver N, Quee PJ, Boos HB, Simons CJ, de Haan L; GROUP investigators. Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *Int J Methods Psychiatr Res.* 2012;21(3):205–221.
9. Gülöksüz S, Pries L-K, Delespaul P, et al. Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry*. 2019;18(2):173–182.
10. Belbasis L, Köhler CA, Stefanis N, et al. Risk factors and peripheral biomarkers for schizophrenia spectrum disorders: an umbrella review of meta-analyses. *Acta Psychiatr Scand.* 2018;137(2):88–97.
11. Linszen MM, Brouwer RM, Heringa SM, Sommer IE. Increased risk of psychosis in patients with hearing impairment: review and meta-analyses. *Neurosci Biobehav Rev.* 2016;62:1–20.
12. Varese F, Smeets F, Drukker M, et al. Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective- and cross-sectional cohort studies. *Schizophr Bull.* 2012;38(4):661–671.
13. Kraan T, Velthorst E, Koenders L, et al. Cannabis use and transition to psychosis in individuals at ultra-high risk: review and meta-analysis. *Psychol Med.* 2016;46(4):673–681.
14. Davies G, Welham J, Chant D, Torrey EF, McGrath J. A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr Bull.* 2003;29(3):587–593.
15. Padmanabhan JL, Shah JL, Tandon N, Keshavan MS. The "polyenviromic risk score": aggregating environmental risk factors predicts conversion to psychosis in familial high-risk subjects. *Schizophr Res.* 2017;181:17–22.
16. Dong M, Anda RF, Felitti VJ, et al. The interrelatedness of multiple forms of childhood abuse, neglect, and household dysfunction. *Child Abuse Negl.* 2004;28(7):771–784.
17. Finkelhor D, Ormrod RK, Turner HA. Poly-victimization: a neglected component in child victimization. *Child Abuse Negl.* 2007;31(1):7–26.
18. Green JG, McLaughlin KA, Berglund PA, et al. Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication I: associations with first onset of DSM-IV disorders. *Arch Gen Psychiatry.* 2010;67(2):113 123.
19. Curtis D. Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatr Genet.* 2018;28(5):85–89.
20. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22.
21. Hand DJ. Classifier technology and the illusion of progress. *Stat Sci.* 2006;21:1–14.
22. Pries LK, Guloksuz S, Ten Have M, et al. Evidence that environmental and familial risks for psychosis additively impact a multidimensional subthreshold psychosis syndrome. *Schizophr Bull.* 2018;44(4):710–719.
23. Guloksuz S, van Nierop M, Lieb R, van Winkel R, Wittchen HU, van Os J. Evidence that the presence of psychosis in non-psychotic disorder is environment-dependent and mediated by severity of non-psychotic psychopathology. *Psychol Med.* 2015;45(11):2389–2401.
24. Kirkpatrick B, Tek C, Allardyce J, Morrison G, McCreadie RG. Summer birth and deficit schizophrenia in Dumfries and Galloway, southwestern Scotland. *Am J Psychiatry.* 2002;159(8):1382–1387.
25. Messias E, Kirkpatrick B, Bromet E, et al. Summer birth and deficit schizophrenia: a pooled analysis from 6 countries. *Arch Gen Psychiatry.* 2004;61(10):985–989.
26. Pries L-K, Klingenberg B, Menne-Lothmann C, et al. 7.3 Polygenic risk for schizophrenia moderates the influence of childhood adversity on daily-life emotional dysregulation and psychosis proneness. *Schizophr Bull*. 2019;45:98–98.
27. Arango C, Díaz-Caneja CM, McGorry PD, et al. Preventive strategies for mental health. *Lancet Psychiatry.* 2018;5(7):591–604.