ENVIRONMENTAL-RISK (E-RISK) LONGITUDINAL TWIN STUDY
CONCEPT PAPER FORM

Proposing Author: Rachel Latham

Author's affiliation, phone, and e-mail address: SGDP Centre, IoPPN, King's College London, rachel.latham@kcl.ac.uk

Sponsoring Investigator (if the proposing author is a student, a post-doc or a colleague): Helen Fisher

Proposed co-authors: Louise Arseneault, Sean Beevers, Temi Moffitt, Avshalom Caspi, Andrew Beddows, Christian Kieling, Thiago Rocha, Joanne Newbury, Aaron Reuben, Line Rasmussen, Andrea Danese, Valeria Mondelli, Brandon Kohrt

Provisional Paper Title: Childhood exposure to ambient air pollution and the risk of depression in adolescence.

Date: 20 March 2020

Objective of the study and its significance:

Major depression has a lifetime prevalence of 11%[1] and is a leading cause of mental-health related disease burden[2]. It is especially debilitating because it often develops in adolescence[3] and tends to have a chronic course. Knowledge about early risk factors for depression is critical so that we can identify those individuals who are at high risk and intervene early to prevent its onset and a lifetime of suffering.

Addressing this, Rocha and colleagues[4] recently developed a multivariable prognostic model using social and demographic risk factors measured in adolescence (sex, skin colour, drug use, school failure, social isolation, relationship with and between parents, fight involvement, child maltreatment, history of running away from home) to identify those individuals who are at high risk for developing major depressive disorder by age 18. This model was shown to predict, with reasonable accuracy, the future onset of depression within the UK E-Risk cohort[4]. However, its prediction was far from 100% and a substantial proportion of adolescents were not correctly identified as being at risk. This suggests that other factors are also involved and their inclusion in the prognostic model may improve its predictive ability.

One potential factor is exposure to air pollution. This is already considered to be a major cause of physical and health morbidity[5,6] and there is accumulating evidence of a possible link between air pollution exposure in childhood and later mental health problems, including depression. For example, Yolton and colleagues[7] found associations between exposure to traffic-related air pollution during childhood and elevated symptoms of depression and anxiety at age 12. Moreover, in the UK, recent pilot work using a subsample of the E-Risk cohort and high-resolution air pollution estimates found strong and robust associations between estimates of outdoor air pollution (particulate matter, $PM_{2.5,}$ and nitrogen dioxide, NO_2) exposure and age-18 depression diagnosis[8].

However, this focused only on a sample of London-based participants so it is unknown whether similar associations would be found across other urban and rural areas of the UK. Therefore, we will extend this pilot work by utilising air pollution (NO_2, NO_x, $PM_{2.5}$, PM_{10}) exposure data for the full E-Risk cohort sample at age 10 to examine longitudinal links with major depressive disorder at age 18 (Objective 1). Focusing on any significant associations observed, we will then investigate whether the inclusion of these estimates of air pollution in the already-developed UK depression-risk model[4] improves the identification of adolescents who are at greatest risk of developing depression (with the outcome being presence of

depressive disorder at age 18) (Objective 2). Note the findings from these first two objectives will form the basis of a single journal publication.

In addition, to obtain pilot data for a future grant application, we will undertake a preliminary investigation of age-18 inflammation (using the following markers: C-reactive protein (CRP), interleukin 6 (IL-6) and soluble urokinase plasminogen activator receptor (suPAR)) as a potential mediator of any observed associations between childhood exposure to air pollutants and age-18 depression diagnosis (Objective 3). Finally, we will examine correlations between age-18 inflammation and depression risk scores produced by the prognostic model (Objective 4). Understanding mechanisms underlying these associations is essential to inform the development of preventive interventions and, ultimately, reduce any harmful impact of air pollution on adolescent depression.

Currently, very little is known about potential biological mechanisms, but it is plausible to consider the role of inflammation. Inflammation is part of the body's innate immune response to infection and disease and its timely response is a critical defence mechanism. However, chronic inflammation damages the body and is now also known to affect the brain. In particular, evidence suggests a role of inflammation in the aetiology of depression. For example, findings of cross-sectional observational studies have shown small elevated levels of circulating inflammation biomarkers among depressed individuals[9]. Furthermore, experimental studies using animal models have shown that inducing an inflammatory state produces affective symptoms such as fatigue, social withdrawal and altered cognition[10] with similar effects also evident in humans[11] thus suggesting a causal effect of inflammation on depression.

As well as these links between inflammation and depression, there are also links between air pollution and inflammation. Air pollutants such as nitrogen dioxide and particulate matter have potent oxidative properties which are associated with elevated inflammation[12] and thought to contribute to the increased risk of physical health problems such as cardiovascular and respiratory conditions[13,14]. Moreover, particulate matter is fine enough to enter the bloodstream and cross the blood-brain barrier where it may promote inflammatory responses. Thus, air pollution exposure could be associated with depression by increasing inflammation.

Statistical analyses:

1. We will utilize air pollution (NO_2, NO_x, $PM_{2.5}$, PM_{10}) exposure estimates using latitude-longitude coordinates of E-Risk twins' home addresses derived from the CMAQ-Urban model. These data will be securely merged with phenotypic E-Risk data. We will conduct a descriptive check of the correlation between children's air pollution exposure at age 10 and age 18.

2. We will then use binary logistic regression to examine associations between age 10 air pollution exposure and major depressive disorder at age 18. Analyses will be conducted in Stata, models will adjust for the non-independence of twins using the 'CLUSTER' command, and will include child sex, ethnicity, and neighbourhood socio-economic status (SES) as covariates.

3. To test the robustness of associations, we will repeat the binary regression models detailed in step 2 including the following additional important risk factors – family SES, family psychiatric history, childhood physical maltreatment up to age 10, smoking at age 18, and age-10 depressive symptoms.

4. Additionally, we will conduct the following sensitivity analyses:

    a) We will include urbanicity as an additional control variable to account for urban factors correlated with air pollution.

    b) We will conduct extremes analyses to test whether those participants who lived in areas with the highest quartile of air pollution exposure during childhood have an increased likelihood of age-18 depression diagnosis compared to those living in areas with the lowest quartile of air pollution exposure.

5. We will then include the statistically significant air pollutants from step 3 as predictors in the existing depression-risk prediction model (developed by Rocha et al.[4]). This analysis will be

conducted in R. To test whether inclusion of these air pollution exposure estimates improves the identification of adolescents who are at high risk for developing major depressive disorder at age 18, we will:

a) Assess model performance including (i) discrimination (the model's ability to separate adolescents with and without major depressive disorder at age 18) based on the C-statistic, (ii) calibration (the agreement between the observed and predicted outcomes) by examining the intercept and slope of the calibration plot, (iii) the Brier score (a combination of discrimination and calibration), and (iv) R^2 (overall goodness of fit measure).

b) Use net reclassification improvement (NRI) method to assess how much model performance has improved. This method quantifies the extent to which the new prediction model (with pollution exposure included) appropriately classifies individuals into risk categories compared to the original prediction model (without pollution exposure).

6. We will use cross-sectional mediation models to conduct an exploratory analysis investigating age-18 inflammation (combined measure and then each marker separately) as a possible mechanism underlying any significant associations observed in step 3 between estimates of childhood exposure to air pollutants at age 10 and depression diagnosis at age 18. We will additionally control for BMI and body temperature at age 18.

7. Finally, we will examine correlations between the depression risk scores calculated by the prediction model in step 5 and age-18 inflammation (combined measure and each marker separately).

Variables Needed at Which Ages (names and labels):

Study: E-Risk Study

Age 5:

General study variables:	
FAMILYID	Unique family identifier
ATWINID	Twin A ID (ex chkdig)
BTWINID	Twin B ID (ex chkdig)
RORDERP5	Random Twin Order
RISKS	Sample Groups
COHORT	Cohort
ZYGOSITY	Zygosity

Previous depression screening for risk prediction model:	
DEPRSE5	Depression scale elder twin

Risk prediction model predictors and/or logistic regression model covariates:	
SAMPSEX	Sex of Twins: In sample
SETHNIC	Ethnicity of Twins
SESWQ35	Social class composite

Age 7:

Previous depression screening for risk prediction model:	
DEPRSE7	Depression scale elder twin

Age 10:

Pollution exposure estimates:	
LOCATION1_NO2_E_P10	NO2 levels at address location 1

LOCATION1_NOX_E_P10	NOx levels at address location 1
LOCATION1_PM2_5_E_P10	PM2.5 levels address location 1
LOCATION1_PM10_E_P10	PM10 levels at address location 1

Previous depression screening and logistic regression model covariate:	
DEPRSE10	Depression scale elder twin

Covariates:	
HARME510	Child harm phase 5-10 elder twin
P10CACOR	Neighbourhood deprivation at age 10
ONS Urbanicity score – age 10	As used by Jo in Newbury et al. 2019 JAMA Psych paper (pollution & psychotic experiences)

Age 12:

Covariates:	
FHANYPM12	Proportion of family members with valid data who have any disorder

Low IQ screening for risk prediction model:	
IQ12E	Pro-rated IQ (INF & MR), 12E

Previous depression screening for risk prediction model:	
CDIE12	Depression Scale - CDI – Elder

Risk prediction model predictors:	
HARME512	Child harm phase 5-12 – Elder twin
EDUCPRFE12	School performance (English & Maths average) - P12 – Elder
SISOCE12	Social isolation (categorical) - P12 – Elder
con23ec12	Have you ever run away from home and stayed away for the night? – Elder
'Any drug use' variable used in Rocha et al 2020. Comprised from: sub1ec12 - substance use – option 1 (upgrade of items 15, 16, and 17) – Elder sub2ec12 - substance use – option 2 (downgrade of items 03 and 07) – Elder	
'Fights' variable used in Rocha et al 2020. Comprised from: con03ec12 Do you sometimes hit someone when you are having an argument? – Elder con04ec12 Do you sometimes start fights with people? – Elder	

Age 18:

Pollution exposure estimates:	
Location1_PM10_E	PM10 levels at address location 1
Location1_PM2_5_E	PM2.5 levels at address location 1
Location1_NOX_E	NOx levels at address location 1
Location1_NO2_E	NO2 levels at address location 1

Depression outcome:	
DXMDEE18	Major depressive episode, dsm4 – P18 – Elder

Inflammation:	
LCA_inflam_LJHR2019	Inflammation latent classes based on CRP/IL-6/suPAR
CRP_suPAR	Categorical variable stratifying by CRP/suPAR
IL6_suPAR	Categorical variable stratifying by IL-6/suPAR
CRP_IL6	Categorical variable stratifying by CRP/suPAR

CRPE18_4SD	CRP concentration (mg/L) with outliers removed - P18 – Elder
lnCRP_E18_4SD	Log-transformed CRP with outliers removed – P18 – Elder
CRP_high	Categorical variable stratifying by high or low CRP
IL6_E18_4SD	Plasma IL-6 with outliers removed – P18 – Elder
lnIL6_E18_4SD	Log-transformed IL-6 with outliers removed – P18 – Elder
IL6_high	Categorical variable stratifying by high or low IL-6
suPAR_E18_4SD	Plasma suPAR with outliers removed – P18 – Elder
suPAR_high	Categorical variable stratifying by high or low suPAR

Covariates:	
smkdlye18	Ever a daily smoker, elder
SMKCURE18	Smoking daily - current - P18 – Elder
BMIE18	bmie18
BTEMPE18	Body temperature, Celsius - P18 – Elder

References cited:

1. Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the national comorbidity survey-adolescent supplement: prevalence, correlates, and treatment. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(1), 37-44.
2. Herrman, H., Kieling, C., McGorry, P., Horton, R., Sargent, J., & Patel, V. (2019). Reducing the global burden of depression: A Lancet–World Psychiatric Association Commission. *The Lancet, 393*(10189), e42-e43
3. Thapar, A., Collishaw, S., Pine, D. S., & Thapar, A. K. (2012). Depression in adolescence. *The Lancet, 379*(9820), 1056-67.
4. Rocha, T. B., Fisher, H. L., Caye, A., Anselmi, L., Arseneault, L., Barros, F.,… & Kieling, C. (2020). Identifying adolescents at risk for depression: A prediction score performance in cohorts based in three different countries. *Journal of the American Academy of Child & Adolescent Psychiatry,* doi: 10.1016/j.jaac.2019.12.004.
5. World Health Organisation, (2013). Review of Evidence on Health Aspects of Air Pollution – REVIHAAP Project: Final Technical Report. WHO Regional Office for Europe, Copenhagen.
6. Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., … & Feigin, V. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet, 389*(10082), 1907–1918.
7. Yolton, K., Khoury, J. C., Burkle, J., LeMasters, G., Cecil, K., & Ryan, P. (2019). Lifetime exposure to traffic-related air pollution and symptoms of depression and anxiety at age 12 years. *Environmental Research, 173*, 199-206.
8. Roberts, S., Arseneault, L., Barratt, B., Beevers, S., Danese, A., Odgers, C. L., …& Fisher, H. L. (2019). Exploration of NO2 and PM2.5 air pollution and mental health problems using high-resolution data in London-based children from a UK longitudinal cohort study. *Psychiatry Research, 272,* 8-17.
9. Howren, M. B., Lambkin, D. M., & Suls, J. (2009) Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. *Psychosomatic Medicine, 71*, 171-186
10. Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). Froom inflammation to sickness and depression: when the inflammation system subjugates the brain. *Nature Reviews Neuroscience, 9*, 46-56.
11. Musselman, D. L., Lawson, D. H., Gumnick, J. F., Manatunga, A. K., Penna, S., Goodkin, R. S., … & Miller, A. H. (2001). Paroxetine for the prevention of depression induced by high-dose interferon alfa. *The New England Journal of Medicine, 344*, 961-966
12. Pope III, C. A., Bhatnagar, A., McCracken, J. P., Abplanalp, W., Conklin, D. J., & O'Toole, T. (2016). Exposure to fine particulate air pollution is associated with endothelial injury and systemic inflammation. *Circulation Research, 119*, 1204-1214
13. Kelly, F. J. & Fussell, J. C. (2015). Air pollution and public health: emerging hazards and improved understanding of risk. *Environmental Geochemistry & Health, 37*, 631-649.
14. Rückerl, R., Schneider, A., Breitner, S., Cyrys, J., & Peters, A. (2011). Health effects of particulate air pollution: A review of epidemiological evidence. *Inhalation Toxicology, 23*, 555-592

Data Security Agreement

Provisional Paper Title	Childhood exposure to ambient air pollution and the risk of depression in adolescence.
Proposing Author	Rachel Latham
Today's Date	19 March 2020

Please keep one copy for your records
(Please initial your agreement)

_RL_ I am familiar with the King's College London research ethics guidelines (https://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx) and the MRC good research practice guidelines (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/good-research-practice/).

_RL_ My project has ethical approval from my institution.

_RL_ I am familiar with the EU General Data Protection Regulation (https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-3-consent-in-research-and-confidentiality/), and will use the data in a manner compliant with its requirements.

_RL_ My computer is (a) encrypted at the hard drive level, (b) password-protected, (c) configured to lock after 15 minutes of inactivity, AND (d) has an antivirus client which is updated regularly.

_RL_ I will treat all data as "restricted" and store in a secure fashion.

_RL_ I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.

_RL_ I will not merge data from different files or sources, except where approval has been given by the PI.

_RL_ I will not post data online or submit the data file to a journal for them to post.
Some journals are now requesting the data file as part of the manuscript submission process. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to the study PI for strategies for dealing with data sharing requests from Journals.

_RL_ Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.

_RL_ I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.

_RL_ I will delete the data after the project is complete.

_RL_ **For projects using location data:** I will ensure geographical location information, including postcodes or geographical coordinates for the E-Risk study member's homes or schools, is never combined or stored with any other E-Risk data (family or twin-level data)

____ **For projects using genomic data:** I will only use the SNP and/or 450K data in conjunction with the phenotypes that have been approved for use in this project at the concept paper stage.

Signature: RLathaM.