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

Proposing Author: Thiago Botter-Maio Rocha

Author's affiliation, phone, and e-mail address: Hospital de Clinicas de Porto Alegre, RS-Brazil; +555133598413; tbotter@hcpa.edu.br

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

Proposed co-authors: Christian Kieling, Helen Fisher, Louise Arseneault, Renate Houts, Terrie Moffitt, Avshalom Caspi

Provisional Paper Title: Development and validation of a composite risk score to predict depressive disorder among youth.

Date: 26th July 2017

Objective of the study and its significance:

To generate and externally validate a composite risk score based on childhood clinical and sociodemographic characteristics of 12-year-old adolescents, free of previous or present depressive symptoms, to predict their risk of developing a depressive episode at 18 years of age.

Mental and behavioural disorders are responsible for 7.4% of global disability adjusted life years (DALYs).[1] Major depressive disorder (MDD) is a major contributor to chronic disease burden worldwide, either directly – MDD is the second leading cause of years lived with disability (YLD) around the globe, or indirectly – it is associated with an increased risk of developing conditions such as diabetes mellitus, heart disease and stroke. There is also well-documented association between MDD and death by suicide. It is estimated that up to 50% of the 800,000 suicides per year worldwide occur within a depressive episode and patients with MDD are almost 20-fold more likely to die by suicide than the general population. Those striking estimates are a result of its high prevalence, early incidence in life, and chronicity.[2]

Prognostic research has received increasing attention in the last few years.[3] However, while the widespread use of tools that predict clinical outcomes in medical practice has promoted the development and testing of preventive interventions with remarkable benefits for population health, such approaches have rarely been attempted for mental health. Furthermore, external validation of such tools has usually been focused on very similar settings in terms of design and sociocultural attributes, whereas clinical practice has incorporated its use in a much less restricted scale.[4]

We are developing a multivariate risk score to predict MDD in young adulthood based on clinically available, easily assessed, reliable sociodemographic variables in order to predict future depressive episodes within a 3-4 year period. Implications for the field are numerous. The model can be used by clinicians to guide long-term decisions based on the estimated risk of a given child. Also, it provides a framework for testing the effectiveness of preventive interventions focused on high-risk individuals. Furthermore, research can benefit from the score in the study of at-risk individuals in terms of neurobiological features, enhancing the knowledge of the biology behind the risk.

Our group has been working to develop a composite risk score for 15-year-old adolescents, free of

previous or present depressive symptoms, capable of predicting their risk of developing a depressive episode at 18/19 years of age. In order to accomplish that, we have used data derived from the 1993 Pelotas Birth Cohort Study, selecting clinically relevant variables, conceptually associated with depressive disorders. We have parsed them into three groups: adolescent-obtained, parent-obtained, and genetic variables. Our final selection included 17 adolescent-obtained, 9 parent-obtained, and 1 genetic. We have then defined as the standard model the one using only the adolescent-obtained variables, as those were considered more easily accessible in a 'real world' scenario.

We are attempting to replicate this model in the Dunedin Study and would also like to test it in the E-Risk Longitudinal Twin Study to ascertain if it translates to the UK as well. We are aware that E-Risk does not have data available at age 15 so we propose to test our composite risk score among 12-year-old adolescents instead.

Statistical analyses:

In the Pelotas cohort, as our first step, we compared the predictive performance of three algorithm modelling strategies (Support Vector Machine, LASSO and Random Forest) and one data-modelling strategy (Logistic Regression Analysis) in order to select the one with best discriminative capacity. All of them showed similar results, with non-significant differences. As logistic regression analysis has been more frequently used in prognostic research, with greater comparability with other studies' results, we have selected it as our primary analysis strategy for the remaining steps.

In our following step, we evaluated the predictive capacity of our standard model, using only the adolescent-obtained variables. Then, we further tested if the addition of the remaining variable groups could increase the predictive capacity of the model. We have defined "increase" as a statistically significant augmentation in the c-statistic, indicating the benefit of the combined model would be above and beyond the potential obstacles and costs associated with the requirement of an additional informant or genetic testing. The results identified that, from a point prevalence of depressive episode at 18/19 years of 3.3% for the included sample, our standard model produced a C-statistic of 0.759 The inclusion of the additional groups of variables did not significantly increase the AUC of the model.

As validation, either internal or external, of predictive model results is a crucial step in prognosis research, we would like to assess whether our results can be validated in the E-Risk Study sample. As our validation strategy process, once we receive the variables requested, we aim to evaluate the comparability of each variable received with ours, making the necessary adjustments to enhance compatibility of datasets. After that, we intend to use our dataset as the development sample, from where we can derive the regression coefficients of variables, and then test in the E-Risk Study dataset as one of the validation samples.

Variables Needed at Which Ages (names and labels):

Study: E-Risk

Age 5

FAMILYID	Unique family identifier
ATWINID	Twin A ID (ex chkdg)
BTWINID	Twin B ID (ex chkdg)
RORDERP5	Random Twin Order
RISKS	Sample Groups
COHORT	Cohort
SAMPSEX	Sex of Twins: In sample
ZYGOSITY	Zygosity
SESWQ35	Social Class Composite
SETHNIC	Ethnicity of Twins

Age of Mum at twins' birth
Age of Dad at twins' birth

TOTEMOE5	CBCL Total Mum & Teacher Emotional Scale (Ex Somatic) - Elder twin
DEPRSE5	Depression scale (from CBCL) - Elder Twin

Age 7
TOTEMOE7 CBCL Total Mum & Teacher Emotional Scale (Ex Somatic) - Elder twin
DEPRSE7 Depression scale (from CBCL) - Elder Twin

Age 10
TOTEMOE10 CBCL Total Mum & Teacher Emotional Scale (Ex Somatic) - Elder twin
DEPRSE10 Depression scale (from CBCL) - Elder Twin

Age 12
BMIE11 Body Mass Index, Age 11 - Elder
HARME512 Physical maltreatment between ages 5-12 – elder (informant report)
CDIE12 Depression Scale – CDI – Elder Twin (self-report)
CDICATE12 Clinically significant depression (CDI >= 20) - P12 - Elder (self-report)
SISOCE12 Social isolation (low/moderate/high) - P12 – Elder

Individual antisocial behavior items (self-report):
conduct3 - Do you sometimes hit someone when you are having an argument? elder
conduct4 - Do you sometimes start fights with people? elder
conduct23 - Have you run away from home and stayed away for the night? Elder

SUB1EC12 Substance Use - Option 1 (Upgrade of Items 15, 16 and 17) - Elder (self-report)
SUB2EC12 Substance Use - Option 2 (Downgrade of Items 03 and 07) - Elder (self-report)
EDUCPRFE12 School performance (English & Maths average) - P12 - Elder

Age 18
DXMDEE18 Major depressive episode, DSM-IV - elder
MDESXE18 Major depression symptom count - elder
CTQCCTOTE18 CTQ - types of abuse or neglect at none, mild, mod/severe level, grouped 0 ,1-2, 3-5 - P18 – Elder (note this self-report for the period from 0-12 years)
ACTQE18 Childhood Trauma (CTQ) total score adjusted - P18 – Elder (self-report for the period from 0-12 years)

References cited:

1. Vos T, Barber RM, Bell B, Bertozzi-Villa A, Birykov S, Bollinger I, *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 743–800 (2015).

2. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. *Nature Reviews Disease Primers* 2, Article number: 16065 (2016). doi:10.1038/nrdp.2016.65.

3. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med. 2015;162:55-63. doi: 10.7326/M14-0697.

4. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. (2013) Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. PLoS Med 10(2): e1001381. doi:10.1371/journal.pmed.1001381.

Data Security Agreement

Provisional Paper Title	Development and validation of a composite risk score to predict depressive disorder among youth.
Proposing Author	Thiago Botter-Maio Rocha
Today's Date	25th July 2017

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

TR I am current on Human Subjects Training (CITI (www.citiprogram.org) or training in human subject protection through my post or courses.

TR My project is covered by Duke or King's IRB OR I have /will obtain IRB approval from my home institution.

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

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

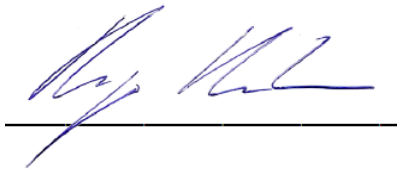
TR I will not post data online or submit the data file to a journal for them to post.
Some journals are now requesting the data file as part of the manuscript submission process. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access.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

TR 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.

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

TR I will return all data files to the Data Manager after the project is complete. Collaborators and graduates of DPPP may not take a data file away from the DPPP office. The data remains the property of the Study and cannot be used for further analyses without express, written permission.

TR 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)

Signature: 

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author: Thiago Botter-Maio Rocha

X I have read the E-Risk data-sharing policy guidelines and agree to follow them

Provisional Paper Title: Development and validation of a composite risk score to predict depressive disorder among youth.

Potential co-authors: Christian Kieling, Helen Fisher, Louise Arseneault, Renate Houts, Terrie Moffitt, Avshalom Caspi

Potential Journals: JAMA Psychiatry, American Journal of Psychiatry

Intended Submission Date (month/year): Autumn 2017

Please keep one copy for your records and return one to Louise (louise.arseneault@kcl.ac.uk)

B. To be completed by potential co-authors:

☐ Approved ☐ Not Approved ☐ Let's discuss, I have concerns

Comments:

Please check your contribution(s) for authorship:

☐ Conceptualizing and designing the longitudinal study

☐ Conceptualizing and collecting one or more variables

☐ Data collection

☐ Conceptualizing and designing this specific paper project

☐ Statistical analyses

☐ Writing

☐ Reviewing manuscript drafts

☐ Final approval before submission for publication

☐ Acknowledgment only, I will not be a co-author

Signature: .....