Concept Paper Template

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

Proposing Author: Thiago Botter-Maio Rocha

Author's Email: tbotter@hcpa.edu.br

P.I. Sponsor: Terrie Moffitt
(if the proposing author is a student or colleague of an original PI)

Today's Date: 20/02/2017

Please describe your proposal in 2-3 pages with sufficient detail for helpful review.

Objective of the study:

To generate and externally validate a composite risk score based on childhood clinical and sociodemographic characteristics of 15 years old adolescents, free of previous or present depressive symptoms, to predict their risk of developing a depressive episode at 18/19 years of age.

Data analysis methods:

Our group has been working to develop a composite risk score for 15 years 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.

As our first step, we compared the predictive performance of three algorithm-modeling strategies (Support Vector Machine, LASSO and Random Forest) and one data-modeling 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 have 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 have 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 a predictive model results is a crucial step in prognosis research, we would like to assess whether our results can be validated in the Dunedin Cohort Study sample.

As our validation strategy process, once we receive the variables requested, we aim to evaluated the comparability of each variable received with ours, making the necessary adjustments to enhance compatibility of datasets. After that, we intent to use our dataset as the development sample, from where we can derive the regression coefficients of variables, and then test in the Dunedin Cohort Study dataset as the validation sample.

Variables needed at which ages:

Age 0 – Perinatal Data: Mother's Age = B1
Age 3: Mother's Age: III-20 B61, Father's Age = III-20 B62
Age 15:

Depression symptoms = XV-12 A48, 55, 59, 61, 62, 84, 109; XV-31 A388-417; B14-21

Gender = XV-23 A209

Ethnicity = XV-122 A57-59; XV-131 A9-11; XV-143 A5

Drugs = XV-39 A440-443; XV-44 A487; XV-62 A1 and A2

Fights = XV-11 A51; XV-40 A445

Run away = XV-39 A424

Friends = XV-27 A241-253; XV-130 A54; XV-135 A41 and A48

Family = XV-7 A6-27; XV-26 A220; XV-27 A229-240; XV-29 A274, A275; XV-130 A53

Body mass index = XV-23 E1 and E2

School failure = XV-26 A217

Childhood maltreatment = variables used in Caspi et al, Science 2002, and Caspi et al, Science 2003.

Age 18: Depression symptoms = XVIII-38 A325-366; B66-68

Significance of the Study (for theory, research methods or clinical practice):

Mental and behavioral 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.

It is also well described the MDD association to 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 been receiving increasing attention in the last 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 approach has been rarely 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 propose the development of a multivariate risk score to predict MDD in young adulthood based on clinically available, easily assessed, reliable socio-demographic variables in order to predict future depressive episodes within 3-4 years 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.

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

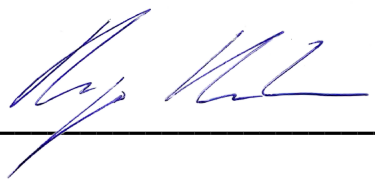
A

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	20/02/2017

Please keep one copy for your records and return one to the PI Sponsor

Please initial your agreement

TR	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
TR	My project is covered by Duke or Otago 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 Dunedin 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.

Signature: 

CONCEPT PAPER RESPONSE FORM

A

Provisional Paper Title	
Proposing Author	
Other Contributors	
Potential Journals	
Today's Date	
Intended Submission Date	

Please keep one copy for your records and return one to the proposing author

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:

_____