E-Risk Study Concept Paper template

Provisional Paper Title: Automating coding of maternal expressed emotion to predict adolescent psychopathology
Proposing Author: Bahman Mirheidari
Author's Email: b.mirheidari@sheffield.ac.uk
Academic supervisor: Click or tap here to enter text. (if the proposing author is a student)
E-Risk Sponsor: Helen Fisher (if the proposing author is not an E-Risk co-investigator)
Today's Date: 10th July 2023
Please indicate if you will require an E-Risk independent reproducibility check: No

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

Background & objective of the study:

Mental health problems affect around a fifth of children and adolescents worldwide (1). Moreover, for the overwhelming majority of individuals, mental illnesses (such as anxiety, depression, schizophrenia, antisocial personality disorder, and bipolar disorder) have their onset in childhood and adolescence with almost three quarters receiving a diagnosis by the time they turn 18 (2). These disorders often have a chronic course affecting individuals for the rest of their lives. The impact of mental illnesses occurs either directly – depression for instance is a leading cause of years lived with disability across the world (3) - or indirectly - through increased risk of developing conditions such as diabetes mellitus, heart disease and stroke in later life (4). The impact of mental disorders is proportionally even larger among youth, a group for whom the benefits of currently available treatments are modest at best (5) and there is scarce mental healthcare provision (6). Therefore, in order to combat this major global health challenge, research is urgently needed to identify youth at the greatest risk of future mental illness as early as possible in order to effectively target preventive interventions to reduce the onset of these difficulties and the associated morbidity and premature mortality (7).

For over five decades, levels of expressed emotion (EE) within families have been studied by psychologists and psychiatrists to determine which adults with mental illness are likely to have the poorest outcomes (8–10). EE refers to the attitudes of caregivers towards their ill relative and comprises criticism, hostility, and/or emotional over-involvement as well as the degree of warmth shown towards them. EE was measured originally through indepth face-to-face interviews (8,10) but due to time constraints has subsequently been assessed through brief samples of caregivers speaking freely about their relative (known as the Five-Minute Speech Sample [FMSS]11). Coding of EE from these easy-to-collect speech samples focuses on the emotions that are apparent from the way in which the caregiver speaks about their relative drawing on both *what is said* and the *tone of voice* used. EE is one of the most robust predictors of poor outcome in the adult mental health literature, with high levels predicting greater likelihood of relapse, poorer clinical course, and worse treatment response across the full spectrum of mental disorders (12,13).

Recently, the focus of EE research has shifted to prediction of mental health problems in young people (14). Within the cohort being utilised for this project, the FMSS method was adapted for use with mothers of young children in the general population and it was demonstrated, using a genetically-sensitive design, that high negativity and low warmth rated from maternal speech samples play a causal role in the development of antisocial behavioural problems in children (15)

and subsequent serious mental illnesses (16). Other studies have shown that ratings of negative emotions from parents' speech predict the onset and course of other mental health problems in children including anxiety (17), depression (18), and attention-deficit hyperactivity disorder (19), underlining its usefulness as an early predictor of youth mental health difficulties. Given that our preliminary epidemiological findings in the E-Risk cohort indicate that high maternal negativity is most robustly associated with all assessed dimensions of psychopathology at age 18 in this sample, we will focus on these in this automated analysis.

However, this promising prediction method is rarely used, because the coding of speech is labour-intensive and requires highly trained raters. Moreover, human ratings have limited reproducibility as they are prone to drift and unconscious biases. Recent developments in computational linguistics make it possible to represent, learn and rate natural speech using automated speech analysis (20,21). Key advancements include bidirectional representation that analyses an utterance in the context of both preceding and following text segments (20), and the application of deep convolutional neural networks to speech recognition and synthesis (21). The most promising approaches involve multimodal processing that integrate these methods to provide a joint representation of speech acoustics, form and content (22,23). Multiple publications have shown that automated analysis of speech obtained from brief audio-recordings can distinguish individuals with mental disorders from unaffected controls (24,25) and accurately determine the severity of their symptoms (26,27). For example, a recent study conducted on speech samples from 157 individuals found that multimodal representations of their speech distinguished those with depression from controls and from those with bipolar disorder with levels of balanced accuracy that were significantly above chance (0.88 and 0.75, respectively), and higher than predictions previously reported in the literature (28). However, these methods have not been applied to predict the future development of mental disorders among adolescents, highlighting an important gap that will be addressed by this project.

It is important to acknowledge that there is a real potential for prejudices and other biases to creep into automated systems through basing the development of models on unrepresentative or skewed data. Indeed, there has been recent controversy surrounding the racially and gender-biased application of facial recognition software (29). Thus, it is crucial for any new automated models to be developed in a responsible manner with as minimal bias as possible. In relation to speech samples, especially when coding the tone and form of words used, it will be important to test that the algorithms developed perform equally well across different socio-economic groups and dialects to reduce the likelihood of biased applications. This equality of model performance will also increase the reproducibility of the models across varying groups and contexts, thereby improving both the scientific replication of the findings and the ease of translation into research and clinical practice.

OBJECTIVES: This project therefore intends to achieve the following objectives:

1. Develop a deep learning model based on the data available within the Environmental Risk (E-Risk) Longitudinal Twin Study to detect negative emotional attitudes of mothers towards their adolescent children from brief samples of mothers' speech collected at age 10.

2. Compare the negativity ratings obtained by the automated model with the gold standard ratings made by highly trained humans.

3. Examine the automated model's performance across socio-economic strata and geographical locations.

4. Investigate the capability of the deep learning model of maternal negativity at age 10 to predict adolescents' mental health outcomes at age 18 years.

Note, we have specifically chosen to use only the age 10 measures of EE as these occur closest to the adolescent period, and we thus strategically chose to focus on this age as the project is funded by the MRC Adolescent Mental Health Research Methodology programme. The recent concept

paper from this group on epidemiological analyses of age-10 maternal EE ratings and mental health outcomes at ages 12 and 18 mirrors this approach and forms the foundation for this automation work.

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

Given the additional improvements afforded by recent developments in computational linguistics, automated speech analysis has the potential to estimate the likelihood that an adolescent will develop mental health problems several years later based on easily obtainable maternal speech samples. This may provide a real advantage to practitioners wishing to effectively target preventive interventions at the most vulnerable adolescents and ultimately reduce incidence rates of mental disorders.

Data analysis methods:

1. Development of deep learning algorithm. We propose to apply state-of-the-art natural language processing methods that utilise bidirectional representations of text and audio characteristics and which have been shown to improve accuracy of numerous natural language processing tasks, as well as speech analysis and generation over previously established standards (20,21). We will combine these recently developed methods in a multimodal deep learning procedure that combines information from text and audio to optimise classification and prediction (28). In the initial model layers, we will use a pre-trained language model (e.g., BERT (20)) to represent text and train machine learning models on extracted audio features, including the rhythm, energy, pitch and modulation of speech (21). In a multimodal fusion layer, we will train (bi-directional) Long Short-Term Memory (LSTM) networks to learn a joint audio-textual representation (28). These models will be trained on two distinct tasks:

i. Classification: the first task is to directly classify speech samples across each dimension of the negativity coding scale, thus predicting final labels for each speech sample. The bimodal fusion (audio-textual) representation will be used to predict negativism that has already been coded by trained human raters in held-out samples of individuals 'unseen' in the model development. This is a (multilabel, multiclass) classification task.

ii. Sequence labelling: the second task is to automate the annotation of sentence level classification of negativism from the speech samples as it is currently carried out by human coders. The bimodal fusion representation will be used to identify and annotate the relevant audio-textual sequences within each sample transcript in the held-out data. Final labels of the negativism coding scale for each sample will then be derived from these identified sequences.

2. Testing the performance of the model. During model development, we will use nested cross-validation so that the majority of our sample is used for training, but the test will remain unbiased. In each iteration, approximately 80% of the total available sample will be used as training, and 20% as testing. We will monitor prediction decrements in testing and validation samples as learning curves to diagnose underfitting and avoid overfitting in predictive models. The performance of the models will be evaluated using f1 and area under the receiver-operating characteristic curve (AUC) to ascertain the accuracy of distinguishing speech samples with and without negative emotional attitude and calibration metrics to assess agreement between observed human-rated negativism and the predicted probabilities of these derived from the automated models.

3. Investigating unbiased application of the model. Speech qualities and the negativity coded from them in this project are likely to depend on accent or dialect, and degree of socio-economic deprivation (among other factors). Our speech samples were only obtained from mothers and thus we cannot explore the impact of gender on the models developed. The corpus of data being utilised in this project represents a diversity of voices across levels of social deprivation and locations. E-

Risk is a nationally-representative sample comprising more than 900 adolescents growing up in the most disadvantaged homes across England and Wales (with a spread of accents and dialects), along with almost 700 age peers from comfortably off backgrounds, and nearly 600 from wealthier backgrounds. Moreover, it has >93% participation at every wave and this high participation has resulted in an unbiased sample that continues to be representative of the UK population (30,31). We will specifically test whether the deep learning models developed perform similarly in terms of identifying negativity across strata of socio-economic deprivation (low vs medium vs high family socioeconomic status; and separately hard-pressed and moderate means vs comfortably off and urban prosperity vs wealthy achiever) on ACORN neighbourhood economic status) and geographical locations (as a proxy for different accents).

4. Prediction of youth mental health outcomes. For comparability with our epidemiological analyses using human ratings, we will first conduct linear regressions to test associations between automated ratings of maternal negativity and each mental health dimensional outcome variable at age 18 in turn, first unadjusted and then adjusted for adolescents' biological sex, their emotional and behavioural problems at age 5 (to attempt to control for reverse causality), and family SES. Family clustering will be accounted for in all of these analyses using robust standard errors with the Huber-White variance estimator.

We will also explore how accurately the automated models of coding negativity from mothers' speech can predict which adolescents will have higher scores on dimensions of psychopathology at 18 years of age (P-factor, internalising, externalising, and thought disorder dimensions) using the root mean square error (RMSE) and mean absolute error (MAE). Analyses will be adjusted for the non-independence of twin observations within families using the Huber-White variance estimator as well as adolescents' biological sex, their emotional and behavioural problems at age 5, highest maternal educational status, and family SES.

Variables needed and at which ages:

Age 5

familyid	Unique family identifier
atwinid	Twin A ID (ex chkdg)
btwinid	Twin B ID (ex chkdg)
rorderp5	Random Twin Order
sampsex	Sex of Twins: In sample
zygosity	Zygosity
seswq35	Social class composite
TOTEXTE5	Total Mum & Teacher Externalising Scale - Elder twin
TOTEMOE5	Total Mum & Teacher Emotional Scale (Ex Somatic) - Elder twin

Age 7

HIEDGM57	Highest Educational Qualification (grouped - mother) P5-P7 Combined

Age 10

Digitised FMSS maternal speech recordings (plus transcripts once available)
Area-Code for where twins were living at age 10 (split into 1-13 categories for different regions if possible)

DISSE10	Dissatisfaction/Negativity towards elder twin
P10CACORNCategory	Acorn Category at Age 10 based on 2001 CENSUS

Age 18

ph_e	P-factor, hierarchical, age 18
intcf_e	Internalizing, 3-factor, age 18
extcf_e	Externalizing, 3-factor, age 18
thdcf_e	Thought disorder, 3-factor, age 18

References cited:

1) Kieling (2011) Lancet;378(9801):1515-25.
2) Kim-Cohen (2003). Arch Gen Psychiatry;60(7):709-17.
3) GBD 2017. Lancet.;392(10159):1789-1858.
4) Prince (2007). Lancet;370(9590):859-77.
5) Cipriani (2016) Lancet;388(10047):881-90.
6) Rocha (2015) Curr Opin Psychiatry;28(4):330-5.
7) Collins (2011) Nature;475:27-30.
8) Rutter (1966) Soc Psychiatry;1:38-53.
9) Brown (1972) Br J Psychiatry;121(562):241–58.
10) Vaughn (1976) Br J Soc Clin Psychiatry;15:157–65.
11) Magana (1986) Psychiatry Res;17:203-12.
12) Butzlaff (1998) Arch Gen Psychiatry;55:547– 52.
13) Hooley (2007) Annu Rev Clin Psychol 3:329–52.
14) Sher-Censor (2015). Dev Rev;36:127-55.
15) Caspi (2004) Dev Psychol.;40(2):149-61.
16) Moffitt (2002) Dev Psychopathol;14:179–206.
17) Gar (2008) Behav Res Ther;46,1266–74.
18) Schwartz (1990) J Psychiatr Res;24(3):231-50.
19) Peris (2003) J Child Psychol Psychiatry 44:1177–90.
20) Devlin (2018) http://arxiv org/abs/1810 04805.
21) van den Oord (2016) https://arxiv org/pdf/1609 03499.
22) Li (2013) IEEE Conference on Affective Computing & Intelligent Interaction;312-7.
23) Palaskar (2018) Proc IEEE Int Conf Acoust Speech Signal Process; http://arxiv.org/abs/1804.09713
24) Ooi (2013) IEEE Trans Biomed Eng;60:497-506.
25) Stasak (2019) Comput Speech & Lang.;53:140-55.
26) Harati (2018) Conf Proc IEEE Eng Med Biol Soc 2018; 2018: 5763-5766.
27) He (2018) J Biomed Inform.;83:103-11.
28) Naderi (2019) KDD;19:doi:10.1145/1122445.1122456
29) Buolamwini (2018) Proc Mach Learn Res;81:1-15.
30) Moffitt &E-Risk Study Team (2002) J Child Psychol Psychiatry;43(6):727-742.
31) Odgers (2012). Dev Psychopathol;24(3):705-21.