University of Glasgow

Investigating temporal and prosodic markers in clinical high-risk for psychosis participants using automated acoustic analysis.

Bianca Bianciardi
MA (Hons)

Thesis submitted in fulfilment of requirements for the Degree of MSc in Psychology (Research)

School of Psychology
College of Medical, Veterinary and Life Sciences
University of Glasgow

September 2022

Abstract

Introduction: Research into language abnormalities has gained attention given the role of language impairments as a plausible marker for early detection and diagnosis of psychosis. Semantic and syntactic aberrations have been widely observed in schizophrenia across illness stages. Recently, acoustic abnormalities such as temporal and prosodic features of speech have been observed in schizophrenia patients. Yet, mixed evidence exists on the presence of acoustic deficits in participants meeting clinical high-risk for psychosis (CHR-P) criteria. The present study aimed to clarify whether acoustic impairments could be used to identify CHR-P individuals when compared to participants with substance use and affective disorders (clinical high-risk negative; (CHR-N) and to healthy controls (HC) participants. Crucially, methodological issues were addressed including the duration of speech samples to determine their impact on the acoustic results.

Methods: Data were available from the Youth mental health, risk and Resilience (YouR) study. Speech samples were recorded from the semi-structured clinical interviews of the Comprehensive Assessment of At Risk Mental States (CAARMS) in 50 CHR-P participants who were compared against a group of 17 HC and 23 CHR-N participants. Temporal and prosodic features were extracted from the recordings. Linear regression was used to determine the influence of interview duration on the acoustic estimates. After examining group differences for each of the acoustic features, temporal and prosodic indices were used to determine whether they could be used determine group status using binary logistic regressions.

Results: No deficits were observed in temporal or prosodic variables in the CHR-P group when compared to HCs. Instead, CHR-N individuals were characterized by slower speech rate, more and longer pauses and higher unvoiced frames percentage compared to CHR-P participants. Temporal features could better discriminate between groups compared to prosodic features, with models explaining up to 47% of the variance between CHR-Ns and HCs and up to 28% of variance between CHR-Ps and CHR-Ns. Yet, none of these models survived bootstrapping. Moreover, group differences for temporal and prosodic features were largely robust to the interview duration effects. Finally, no significant relationship was obtained for temporal and prosodic features with clinical and functional symptom severity.

Discussion: These finding suggests that temporal and prosodic features of speech are not impaired in early-stage psychosis. The acoustic features examined indicated the presence of acoustic impairments in CHR-N participants, which resulted spurious following bootstrapping

and therefore hinted to the importance of employing validation methods on acoustic signatures in psychosis. This is crucial given the small sample sizes across the literature and heterogeneity of the clinical groups. Given the absence of acoustic disturbances of speech in CHR-P individuals observed in the present research, sematic and syntactic abnormalities may constitute a more promising biomarker of early psychosis. Further studies are required to clarify whether acoustic abnormalities are present in sub-groups of CHR-P participants with elevated psychosis-risk.

Table of Contents

List of Tables

List of Figures

Acknowledgements

With sincere gratitude I would like to thank my supervisor, Dr Alessio Fracasso, for his invaluable support and guidance. Thanks for always keeping your virtual and later on (as Covid allowed) physical door open and making me feel that I could always ask for help and advice.

Special thanks to Professor Peter Uhlhaas, for his feedback, guidance and opportunities given throughout my academic journey.

I would also like to extend my thanks to the entire team that previously worked on the MRC funded YouR-Study and collected the data and to the participants who took part in the study, as it was thanks to them that my research was made possible.

I am also grateful to Alessio Fracasso's and Peter Uhlhaas' team members for the helpful feedback during the lab presentations and for the supporting environment.

Finally, a special thanks goes to my parents, Lucia and Piero, for always supporting my ambitions and for their encouragement throughout my degree.

Authors Declaration

Name: Bianca Bianciardi

" I certify that the thesis presented here for examination for a MSc degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice. The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgement. I declare that the thesis does not include work forming part of a thesis presented successfully for another degree. I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research. I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues."

Abbreviations

ADMs	Antidepressant Medications
AIC	Akaike Information Criterion
AMPA	A-amino-3-hydroxy-5-methyl- 4-isoxazolepropionic acid
apq5/ppq5	Five-point Amplitude Perturbation Quotient/five-point Period Perturbation Quotient
APS	Attenuated Psychotic Symptoms
ARMS	At-Risk Mental State
AUC	Area Under the Curve
BACS	Brief Assessment of Cognition in Schizophrenia
BDNF	Brain-derived Neurotrophic Factor
BLIPS	Brief limited intermittent psychotic symptoms
BS	Basic Symptoms
BSABS	Bonn Scale for the Assessment of Basic Symptoms
BSIP	Basel Screening Instrument for Psychosis
CAARMS	Comprehensive Assessment of At-Risk Mental States
CCNi	Centre for Cognitive Neuroimaging
CHR-N	Clinical high-risk Negative
CHR-NT	Clinical High-Risk who will not transition to psychosis
CHR-P	Clinical High-Risk for Psychosis
CI	Confidence Intervals
CNR2	Cannabinoids Receptor 2
CNV	Copy Number Variation
COGDIS/COOPER	Cognitive Disturbances scale/Cognitive-Perceptive Basic Symptoms
DNA	Deoxyribonucleic acid

DS	Disorganised Speech
DSM	Diagnostic and Statistical Manual of Mental Disorders
DTNBP1	Dysbindin 1
E/I	Excitation/Inhibition
F0	Fundamental Frequency
F1/F2	First Formant/Second Formant
FEP	First Episode Psychosis
FTD	Formal Thought Disorder
GABA	Gamma-aminobutyric acid
GAF	Global Assessment of Functioning
GF	Global Functioning
GRD	Genetic Risk and Deterioration syndrome
GWAS	Genome-wide association studies
HC	Healthy Controls
HG	Heschl's gyrus
HNR/NHR	Harmonics to noise ratio/Noise to harmonics ratio
IRRs	Incidence Rates Ratio
K-FTDS	Kiddie Formal Thought Disorder Rating Scale
LSA	Latent Semantic Analysis
MINI	Mini-International Neuropsychiatric Interview
MRC	Medical Research Council
NBI	Non-Bizarre Ideas
NMDA	kainate and N-methyl-D-aspartate
NR2	Nagelkerke pseudo R^2

OR	Odds Ratio
PA	Perceptual Abnormalities
PANSS	Positive and Negative Syndrome Scale
PCA	Perceptual-Cognitive Anomalies
PCP	Phencyclidin
PFC	Prefrontal Cortex
PLEs	Psychosis-like Experiences
POS	Part-of-Speech
PQ	Prodromal Questionnaire
PTD/NTD	Positive Thought Disorder/Negative Thought Disorder
ROC	Receiver-Operating Characteristic
ScZ	Schizophrenia
SIPS/ SOPS	Structured Interview for Prodromal Syndromes/Scale of Prodromal Symptoms
SN	Substantia Nigra
SPET	Single-photon emission tomography
SPI-CY	Schizophrenia Proneness Instrument, Child and Youth version
SPI-A	Schizophrenia Proneness Instrument, Adult version
ST	Semitones
UHR	Ultra-High Risk
UTC	Unusual Thought Content
VIF	Variance Inflation Factor
VTA	Ventral Tegmental Area
WM	White Matter

YouR-study	Youth Mental Health Risk and Resilience study

1 Introduction

1.1 Schizophrenia - definition and prevalence

Schizophrenia (ScZ) consists of one of the most impairing psychotic disorders, being one of the main public health problems that psychiatry faces affecting nearly 1% of the population worldwide (Saha *et al.*, 2005; World Health Organization, 2013). The disorder typically develops during adolescence, often preceded by a period of subthreshold symptoms defined as clinical high risk for psychosis (CHR-P; McGorry *et al.*, 2018). CHR-Ps criteria consist of subthreshold psychotic experiences or genetic risk and functional decline (Fusar-Poli *et al.*, 2015) as well as self-experienced basic-symptoms (Schultze-Lutter *et al.*, 2016). Yet not all CHR-Ps will transition to psychosis with present data suggesting that around 22% of CHR-Ps will transition to first-episode psychosis (FEP) after 3 years (Fusar-Poli *et al.*, 2020). The onset of ScZ commonly occurs around 5 years earlier in males than in females (Castle, 2000) and is more likely to develop in individuals born in urban areas and is more prevalent in migrants, according to a meta-analysis that examined systematic reviews published between 1965 and 2002 (McGrath *et al.*, 2008). Life expectancy in patients is reduced by 15-20 years compared to the average lifespan in the general population (Laursen, Nordentoft and Mortensen, 2014).

1.2 History of schizophrenia.

In 1911, Eugen Bleuler, a Swiss psychiatrist, coined the term schizophrenia (Bleuler, 1911), which consists of two Greek words skhizein (to split) and phrēn (mind), referring to the fragmentation of psychological functioning that he observed in patients (Fusar-Poli and Politi, 2008). Earlier, a German psychiatrist, Emil Kraepelin had described a distinct psychotic condition, different from other forms of psychosis, and named it “dementia praecox” (Kraepelin, 1896), which Bleuler challenged by calling it neither dementia, nor something precocious. Yet, Bleuler later revisited the definition after close study of his patients and considered the definition of dementia praecox to consist of a cluster of disorders which were not always uncurable, which did not always emerge during adolescence and may not have progressed onto dementia (Fusar-Poli and Politi, 2008). The pioneering works of Kraeplin and, especially Bleuler, presented for the first time a diagnosis of mental illnesses based on psychology, which later influenced the symptoms-based classification of psychiatric disorders leading to the 1st edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-I) (Andreasen, 1989).

1.3 Symptomatology of schizophrenia.

Consisting of a complex mental disorder, ScZ presents a spectrum of symptoms that can be clustered under three main domains: positive, negative and cognitive symptoms (Kay, Fiszbein and Opler, 1987).

1.3.1 Positive symptoms

Positive symptoms can be seen as alterations from normal functioning such as aberrations in perception, thinking and feelings of unreality, nihilism. Positive symptoms are the most overt symptoms of the disorder. Positive symptoms include hallucinations (the misattribution of perceptual experiences to the outside environment), delusions and disorganised speech (inability to produce coherent speech including derailment and tangentiality). These symptoms are generally transient and are present during the acute phases of the disorder (Kay, Fiszbein and Opler, 1987). Hallucinations can be present in any of the 5 senses in ScZ i.e. visual, auditory, olfactory, gustatory and tactile. Yet, auditory hallucinations are the most common type, followed by visual ones whereas hallucinations in the remaining senses are less frequent (Bauer *et al.*, 2011). Crucially, most individuals with psychosis present multi-modal hallucinations rather than experiencing hallucinations in only one sensory domain (Lim *et al.*, 2016). Delusions consist of ideas held with full conviction, even in lack of or with contradicting evidence (Bentall *et al.*, 2001).

The most common types of delusions in psychosis are of paranoid or persecutory nature, although other delusions exist including grandiose, somatic or ideas of reference (where personal significance is attributed to events that are neutral in content). Delusion is considered the most common symptom of psychosis (Appelbaum, Robbins and Roth, 1999) and is highly prevalent in first-episode schizophrenia, reported in 70% of participants (Coid *et al.*, 2013). Formal thought disorder (FTD) indicates a disorder in the organisation and expression of language and in the maintenance of coherent speech rather than being a disorder of language per se (Yalincetin *et al.*, 2017) and is characterised by symptoms including tangentiality, derailment (sudden loss of association in what is being said), distractibility and poverty of speech. FTD will be discussed in more depth in *Section 1.19* of the *Introduction* chapter.

1.3.2 Negative Symptoms

Negative symptoms refer to symptoms that are present in normal experiences but are diminished (McGurk *et al.*, 2004) including apathy/avolition (lack of motivation), anhedonia (lack of pleasure), affective flattening and social withdrawal. Negative symptoms have been observed since the early investigations in ScZ, yet these symptoms have been differently

clustered over the years. Recently, a meta-analysis has subdivided negative symptoms using factor-analysis into two subcategories: 1) amotivation and 2) diminished expression (Foussias *et al.*, 2014). The cluster of amotivation includes avolition, anhedonia social withdrawal and apathy, whereas the diminished expression category includes flattened emotions, unchanged facial expressions, flattened pitch and reduced voice strength and tone and poverty of speech.

1.3.3 Cognitive symptoms

Cognitive symptoms refer to deficits in a wide range of cognitive domains and include impaired attention, reduced verbal fluency, poor abstract thinking, lower psychomotor speed, reduced executive functions and working memory (McGurk and Mueser, 2004; Luck and Gold, 2008). Not all ScZ patients present all these cognitive deficits, with heterogeneous manifestations of cognitive symptoms which are however significantly more stable throughout the course of illness compared to positive or negative symptoms (Harvey *et al.*, 2006). Cognitive deficits are observed in those who do not transition to psychosis, they have been observed in FEPs (Mesholam-Gately *et al.*, 2009) and in unaffected first-degree relatives of psychosis patients (Agnew-Blais and Seidman, 2013; Bora and Murray, 2014). Considering the importance of cognitive symptoms in ScZ, the authors of the DSM-5, considered classifying cognitive deficits as a separate category alongside positive and negative symptoms. However, cognitive impairments remain under negative symptoms in the DSM-5 due to the heterogeneity and the lack of clear-cut specificity of cognitive symptoms in psychosis (Ebenezer, 2015). Cognitive impairments in ScZ contribute to deficits at the functional and intellectual level rendering the cognitive symptoms potentially important for the treatment of the disorder (Green, Horan and Lee, 2015). Yet, cognitive deficits emerge in the absence of symptoms, do not respond to antipsychotic medication (Bowie and Harvey, 2006) and their emergence precedes the onset of psychosis illness (Fusar-Poli *et al.*, 2012; Kahn and Keefe, 2013). For this reason, there is increasing clinical and research interest in the implementation of methods to improve cognition and ultimately improve functioning in ScZ. Cognitive training consists of a promising method for the improvement of cognition in ScZ; metanalytical evidence reported moderate effect sizes of cognitive training on alleviating cognitive and functional deficits and a small effect size on psychotic symptoms (McGurk *et al.*, 2007; Wykes *et al.*, 2011). Nonetheless, cognitive training has proven most effective when combined with psychosocial rehabilitation and despite the promising results, cognitive training presents several limitations including being time-consuming, limited accessibility and being relatively expensive (McGurk *et al.*, 2007; Wykes *et al.*, 2011).

1.4. Aetiology of the disorder.

The aetiology of ScZ comprises genetic and environmental factors. The neurodevelopmental hypothesis of ScZ proposed that prenatal and early life events are the cause of ScZ (McGrath *et al.*, 2003). It is nowadays acknowledged that genetic, prenatal and perinatal risk factors have an important role in later development of psychosis (Jablensky, McNeil and Morgan, 2017) and the interaction between genetic and environmental aspects have shown to be associated with later development of ScZ (Misiak *et al.*, 2018).

One of the most influential neurodevelopmental hypotheses consists in the *hit model* of ScZ (McGrath *et al.*, 2003). This model suggests that genetic and pre/peri-natal experiences lead to neural alterations, whilst later risk factors including excessive synaptic pruning or substance misuse consists of a second hit, which can occur at any time, yet with a lesser effect when occurring later on in life (Pantelis *et al.*, 2003). A more recent multi-hit threshold model considers not only the genetic *hits* but also the environmental factors that act in combination accounting for the complex interactions that lead to the development of ScZ (Davis *et al.*, 2016).

1.4.1 Genetic factors.

An essential aspect of the genetic factors at the basis of neurodevelopmental theories of the disorder suggests that a large number of genes associated with the predisposition can affect the neural development, maturation and differentiation of nerve cells (Henriksen, Nordgaard and Jansson, 2017).

1.4.1.1 Pre-molecular genetics.

Family and twin studies have detected an 80% heritability of the disorder (Shih *et al.*, 2011). Although the incidence of ScZ is of 1% of the general population as stated above, first-degree relatives of patients have a higher probability of developing the disorder which is 8-fold higher and in the presence of two first-degree relatives, the probability increases to 11-fold (Lo *et al.*, 2020). In addition, twin studies have shown a 50% overall probability to develop ScZ if the other monozygotic twin presents the disorder, also, the genetic heritability of ScZ in monozygotic twin studies is 80-85% (i.e., genetic contribution explains 80% of the 50% overall likelihood) (Cardno and Gottesman, 2000). Fraternal twins have instead been shown to present

only a 0-28 % incidence of developing ScZ if the other twin develops the disorder. Although the precise genetic causes of ScZ are still ill-defined, given that there seems to be no single allele/locus associated with ScZ, the family and twin studies hint to a strong genetic component associated with the disorder.

1.4.1.1 Molecular genetics.

It has been shown that the genetic architecture of ScZ is complex and multifactorial and presents several risk variants with low effect sizes and very few with large effect sizes. This has been shown by genome-wide association studies (GWAS; http://www.genome.gov/gwastudies). One of the most influential and ground-breaking large-scale GWAS study , included 37 thousand ScZ patients and 113 thousand healthy controls (HC) and identified 128 ScZ associations with more than 108 risk loci (Ripke *et al.*, 2014). The identified genes included the dopaminergic receptor D2 gene, the genes of the glutamatergic system and tissue plasticity and the genes of calcium channel subunits (Martel and Gatti McArthur, 2020). Yet, each of these gene contributed solely to less than 1% of the development of psychosis (Purcell *et al.,* 2009).

Nonetheless, the 22q11.2 deletion syndrome has been found to increase the risk of developing SZ by approximately 25% (Cleynen *et al.*, 2021). Taken together, the evidence on polygenic risks in ScZ hints at the genetic complexity of the disorder involving multiple risk factors.

Studies investigating *de novo (i.e. new and not genetically inherited)* Copy number variation (CNV) have shown duplication or deletion of genomic sequences in specific regions of deoxyribonucleic acid (DNA) (<1%; Thapar and Cooper, 2013). It was observed that individuals with ScZ compared to controls presented an increase of 1.15-fold of large (>100 kilobase), rare (< 1% in the population) CNVs (International Schizophrenia Consortium, 2008). Importantly, a strong relationship between the large), rare CNVs (1q21.1, Neurexin 1 (NRXN1), 3q29, 15q13.3 and 22q11.2 and duplicates at 16p13.1 and 16p11.2) and ScZ have shown strong associations, acting on multiple genes critical for neural development, cell signalling and glutamate neurotransmission (Xu *et al.*, 2008; 4 *et al.*, 2009; Ripke *et al.*, 2014; Chang *et al.*, 2016; Ruderfer *et al.*, 2016). Yet, the association presented particularly high odd ratios (OR), which may indicate that the as current measures of CNV lag behind GWAS study techniques, suggesting that less frequent CNVs that have not been identified in the GWAS

studies that could potentially play a central role in the development of ScZ (Avramopoulos, 2018).

1.4.2 Environmental factors

There are several important environmental factors that increase the chances of developing ScZ, accounting for about 0-20% risk for schizophrenia (Cardno and Gottesman, 2000). Factors found to be associated with psychosis include but are not limited to obstetric complications (Cannon *et al.*, 2002), prenatal infections, bullying and childhood maltreatment (Varese *et al.*, 2012), migration , and cannabis use (Shih *et al.*, 2011). These exposures are heritable and association with psychosis is in part attributed to genetic influences. For example, from twin studies, it was observed that the association between psychotic experiences and bullying/life events/tobacco were explained by genetics, indicating that these associations are not casual (Dean and Murray, 2022).

1.4.2.1 Prenatal and perinatal risk factors

Several prenatal factors can influence the risk of developing psychosis. Seasonal factors during gestation have shown to influence the risk of psychosis development, with winter and spring seasons being positively correlated (OR 1.07) with a risk of 3.3% of developing psychosis (Davies *et al.*, 2003). Additionally, the location of birth has a minor impact on ScZ risk, whereby children born in urban rather than rural areas present a higher risk of psychosis of 11.73% (Sørensen *et al.*, 2014).

An additional risk factor consists of the father's age, with very small effect sizes for fathers younger than 25 years of age, a small effect size for those aged 35 or over and a medium effect size for mature fathers (≥ 50 years old, when compared to fathers between ages of 25 – 29 years old; Miller *et al.*, 2011). Intriguingly, no association was found with maternal age.

Maternal illness including influenza, rubella, herpes and certain microbial agents have shown mixed effect sizes but have been identified as potential risk factors in ScZ with a 1.32-fold increased ScZ risk in the new-born even after accounting for parental history of psychiatric admission and urbanicity (Nielsen, Meyer and Mortensen, 2016). Evidence has also suggested that maternal malnutrition may play a role, such that iron or other vitamin deficiencies during

pregnancy are associated with a four-fold increase in the child's risk of developing ScZ (Susser, Hoek and Brown, 1998; McGrath *et al.*, 2008).

Obstetric complications, seem to have a modest yet present impact on later ScZ development and include three main group of possible birth complications with an ORs of 1.69 – 7.75 (Cannon, Jones and Murray, 2002). As outlined from the meta-analysis the complications included pregnancy complications (bleeding, diabetes, preeclampsia and blood issues); abnormal foetal development (low birth weight, congenital malformations and small head circumference); delivery complications (asphyxia, uterine atony and emergency caesarean section). Obstetric complications seem to be more strongly associated to ScZ in those individuals with an early illness onset (Rosso *et al.*, 2000).

1.4.3 Later environmental risk factors

1.4.3.1 Substance abuse

One of the main risk factors for later development consists of substance misuse including cannabis, alcohol, hallucinogens, sedatives and abuse of other substances (Nielsen *et al.*, 2017). Cannabis and alcohol consist of the highest risk factors with a 5- and 3-fold increase respectively. Moreover, Cannabis correlated with psychotic symptom development (OR:1.41) and ScZ (OR: 1.82; Moore *et al.*, 2007). The relationship with cannabis is dose-dependent and shows an association with earlier psychosis onset (up to 2.7 years; Donoghue, Golowich and Holstein, 2014). Finally, an endocannabinoid system for the emergence of psychotic experiences has been identified from a recent GWAS with a link between the CNR2 (cannabinoids receptor 2) locus and psychotic experiences (Legge *et al.*, 2019).

1.4.3.2 Childhood adversity

Abuse and neglect, peer bullying, parental loss or divorce and poverty are childhood adversities that strongly impact the risk of developing ScZ (Seidenfaden *et al.*, 2017). Childhood abuse has been shown to have an OR ranging from 1.7 to 15, alongside to other moderating variables such as gender, cannabis use, trauma and depression (Sideli *et al.*, 2012). Crucially, ceasing childhood traumas such as bullying and physical assault reduces the presentation of psychotic experiences as shown by a study conducted in a cohort of 1,112 adolescents (Kelleher *et al.*, 2013).

Furthermore, childhood abuse has been linked to the severity of psychotic symptomatology, whereby females who underwent sexual abuse are more likely to present auditory hallucinations (Misiak *et al.*, 2016). Plausible mechanisms at the basis of the correlation between childhood adversities and psychosis risk include increased inflammation and metabolic dysregulations as well as reduced levels of brain-derived neurotrophic factor (BDNF; Misiak *et al.*, 2016).

1.4.3.3 Migration

First and second-generation migrants have shown to have increased risk of later developing schizophrenia with 2.3 – 2.7 incidence rates ratio (IRRs) and 2.1 – 4.5 IRRs respectively compared to other ethnic groups ; Tortelli et al., 2015). In a study conducted in the United States, a 2-fold increase in the risk of later diagnosis was found in Norwegian migrants compared to native-born Americans (Ødegaard, 1932). Plausible reasons for migration being a risk factor include postmigration aspects such as socioeconomic status, age, gender, discrimination, social isolation, trauma and abuse (Hollander *et al.*, 2016).

1.5 Pathophysiology of schizophrenia.

In the last few decades, a large body of research has aimed to elucidate the neurochemical and brain processes associated with the disorder, that alterations in the neurotransmitter signalling systems are implicated in the manifestation of ScZ. The most well-known pathophysiological hypotheses involve dopamine and glutamate.

1.5.1 The dopamine hypothesis

Being the dominant hypothesis of ScZ for decades, the dopamine hypothesis postulates a dysregulation in the dopamine signalling. The theory is based on the evidence that antipsychotic drugs that reduced positive symptoms in ScZ increased dopamine metabolism (Howes, McCutcheon and Stone, 2015).

Dopamine consists of a key neurotransmitter with a crucial role in the brain for the regulation of essential functions such as motor control, reward and crucial cognitive functions such as executive functions and memory (Robbins and Arnsten, 2009; Bromberg-Martin, Matsumoto and Hikosaka, 2010; Ledonne and Mercuri, 2017). Metabotropic dopamine receptors can be subdivided into two subcategories based on their primary action: D1-like excitatory receptors

(including D1 and D5) and primarily inhibitory D2-like receptors (D2, D3, D4) (Hasbi, O'Dowd and George, 2011). Dopamine is synthesised in two brain regions: the substantia nigra (SN) and ventral tegmental area (VTA) and from these, dopamine is projected within the following pathways: the *nigrostriatal pathway* transmits dopamine from the substantia nigra to the striatum, the *mesolimbic pathway* presents dopamine transmission from the VTA to nucleus accumbent and the *mesocortical pathway* transmits dopamine from the VTA to the prefrontal cortex (PFC; Fabiana *et al.*, 2021).

Originally, the dopamine hypothesis focused on the hyperactivity of dopamine as the aetiological basis of psychosis (Lau *et al.*, 2013). Evidence of hyperdopaminergia resulting in positive symptoms emerged from observation that first-generation antipsychotics block D2 receptors reducing positive symptoms whilst amphetamine and methylphenidate that are dopamine agonists present psychotomimetic properties (Howes and Kapur, 2009). Yet, classical criticisms of the dopamine hypothesis indicate poor efficacy of D2 receptor antagonists in reducing the negative and cognitive symptoms (Javitt, 2007; Javitt *et al.*, 2012) as well as evidence that second-generation antipsychotic that have low levels of D2 receptor occupancy are very effective in reducing positive symptoms of chronic ScZ (Lawrence, First and Lieberman, 2015; Li, L Snyder and E Vanover, 2016).

Later refinement of the hypothesis have postulated that alongside the dopamine hyperactivity in the mesolimbic pathway, hypoactivity in the mesocortical pathway creates negative and cognitive symptoms (Fabiana *et al.*, 2021). Evidence of different cortical and subcortical dysregulation of dopamine emerged from post-mortem, lesion and PET studies indicating striatal dopamine increases reflecting positive symptoms but frontal decreases in dopaminergic activity reflecting negative symptoms (Howes and Kapur, 2009).

Overall, despite the importance of dopamine in the pathophysiology of ScZ, due to the limitations outlined above, other neurotransmitters that are equally important in their involvement in ScZ pathophysiology have been studied.

1.5.2 Glutamate hypothesis

Glutamate is the primary excitatory neurotransmitter in the central nervous system (Howes et al., 2015). Glutamatergic neurons that extend from the superficial layer to the subcortical structures make use of about 60-80 % of the total metabolic activity in the human cerebral cortex(Rothman, Hayes and Summons, 2003). These excitatory cells communicate with other neurons through metabotropic and ionotropic glutamate receptors. Crucially, ionotropic

glutamate receptors consist of a-amino-3-hydroxy-5-methyl- 4-isoxazolepropionic acid (AMPA), kainate and N-methyl-D-aspartate receptors (NMDA-R), named after the agonists that activate these receptors (Kew and Kemp, 2005).

The role of glutamate in the pathophysiology of ScZ has been recognised for several decades, suggesting that the emergence of positive, negative and cognitive symptoms arises from the dysregulation of glutamatergic neurotransmission (Olney and Farber, 1995). Specifically, the glutamate hypothesis emerged from the observation that drugs such as Ketamine and phencyclidine (PCP), could induce ScZ-like symptoms in healthy individuals (Thornberg and Saklad, 1996; Olney, Newcomer and Farber, 1999; Moghaddam, 2003; Stone *et al.*, 2008) and aggravate psychotic symptoms in patients (Itil *et al.*, 1967; Cohen and Benedek, 1979). These non-competitive NMDA-R antagonists disrupt glutamate neurotransmission by blocking NMDA-R site.

The first studies that showed the involvement of NMDA-Rs in the pathophysiology of ScZ looked into the effects of NMDA-Rs in such as Ketamine and PCP, in healthy volunteers (Allen and Young, 1978; Krystal *et al.*, 1994). These studies showed that administration of NMDA-R antagonists could transiently mimic ScZ-like positive, negative and cognitive symptoms in healthy volunteers (Luby *et al.*, 1962; Krystal *et al.*, 1994; Javitt *et al.*, 2001).

Gathering further support for this hypothesis, NMDA receptor deficits have been observed in post-mortem brain tissue derived from individuals with schizophrenia (Stone *et al.*, 2008) and in the brain tissue of those living with the disorder using single-photon emission tomography (SPET) imaging (Pilowsky *et al.*, 2006). Another imaging technique, proton magnetic resonance imaging (1H-MRS), has evidenced decreased in vivo glutamate levels in the medial frontal brain region of those with schizophrenia when compared to healthy individuals (Marsman *et al.*, 2013). Finally, recent genetic evidence has also shown that Copy Number Variants that substantially increase susceptibility to ScZ are enriched for NMDA receptors (Pocklington *et al.*, 2015).

1.5.3 The GABA hypothesis.

Gamma-aminobutyric acid (GABA) consists of a further neurotransmitter involved in the pathophysiology of ScZ and is the main inhibitory neurotransmitter in the central nervous system, crucial for neural synchronisation and connection between brain areas (Zhou and Danbolt, 2014). Aberrant neural synchronisation has been associated with cognitive deficits

and perceptual impairments (Uhlhaas *et al.*, 2010). The GABA hypothesis suggests that aberrant GABA signalling causes excitation and inhibition (E/I) imbalance, ultimately leading to the symptoms and cognitive deficits of ScZ (Lewis *et al.*, 2012; Nakazawa *et al.*, 2012). Evidence in support of this hypothesis comes from post-mortem studies that found reduced GAD67 (an enzyme involved in GABA-synthesis) in PFC (Egerton et al., 2017; Tanaka *et al.*, 2008) which is involved in 90% of GABA production. Mixed evidence emerges from animal models, with some studies showing that injection of $GABA_A$ antagonist picrotoxin in rodents recreated circuit deficits observed in ScZ (Wassef, Baker and Kochan, 2003). Yet, picrotoxin activation can also be linked to dopamine activation. Similarly, following the administration of GABA agonists, clinical studies showed a reduction in ScZ symptoms, yet similarly to animal models the interaction of dopamine could not be ruled out, rendering the GABA deficit unclear. Additionally, meta-analytical findings of 16 proton magnetic resonance spectroscopy studies that examined GABA concentrations observed inconsistent patterns across studies (Egerton *et al.*, 2017).

1.5.4 A revised hypothesis on the role of GABA, glutamate and dopamine pathways.

Recent pathophysiological hypotheses have focused on glutamatergic and GABAergic dysregulation, followed by dopaminergic dysfunction as a secondary consequence in the mesolimbic pathway (Schwartz, Sachdeva and Stahl, 2012). This theory offers the opportunity to reconcile findings that reported dysregulation of NMDA and dopamine activity, indicating that an NMDA reduction leads to hyperdopaminergia in the mesolimbic circuit and hyperdopaminergic in the mesocortical circuit (Schwartz, Sachdeva and Stahl, 2012). It has been suggested that the imbalance between glutamate and NMDA interneurons reduced the control of PFC in ScZ (Balu, 2016).

NMDA receptors are central for a range of functions such as neuronal development of synaptic plasticity, learning, and cell integrity (Hashimoto, Murata and Yoshii, 2017) and through its interactions with the glutamate system have been shown to induce psychosis symptoms through effects on the cortical E/I balance (Gonzalez-Burgos and Lewis, 2012). Thus, the balance between cortical excitation and inhibition has been suggested as a neural mechanism at the basis of psychosis (Gonzalez-Burgos and Lewis, 2012; Lisman, 2012).The cognitive and sensory impairments observed in ScZ may emerge from disturbances in E/I balance (Uhlhaas and Singer, 2010; Uhlhaas, 2013). The regular interplay between excitatory glutamate cells

and inhibitory GABAergic interneurons is central for the emergence of neural rhythmic activity (Gonzalez-Burgos, Hashimoto and Lewis, 2010; Carlen *et al.*, 2012).

There is evidence that both GABAergic and glutamatergic circuits are disrupted in psychosis. The implication of aberrant NMDA-R receptor functioning in circuit dysfunctions in ScZ emerges from evidence of a reduction of NR1 subunit in the PFC (Catts *et al.*, 2016), as well as post-mortem studies showing other NMDA-R subunits (Kornhuber *et al.*, 1989; Akbarian *et al.*, 1996; Weickert *et al.*, 2013). Importantly, research has hinted that NMDA-Rs may be specifically reduced on interneurons that are enriched for GAD(67) mRNA, which is a central enzyme for the synthesis of GABA. Furthermore, during neural development, NMDA receptor aberration involves GABAergic interneurons and in turn affect interneuron maturation and the regulation of glutamate and GABA circuit functioning (Nakazawa, Jeevakumar and Nakao, 2017). Therefore, the E/I balance hypothesis of ScZ is in accordance with the glutamate and GABA hypotheses and would better account for the presence of negative symptoms and cognitive deficits that the dopamine hypothesis fails to explain (Howes, McCutcheon and Stone, 2015).

1.6 The Course of ScZ.

The course of ScZ is commonly subdivided into four stages of the disorder: a premorbid, prodromal, psychotic and stable phases (Fusar-Poli *et al.*, 2013). Negative symptoms are generally present during nonpsychotic periods whereas positive symptoms are characteristic of psychotic episodes (Kandel *et al.*, 2000). During the premorbid period decline in cognition, social and motor functioning are observed, hence symptoms nonspecific to ScZ are generally present during this stage.

While genetic and environmental factors play a role in the development of the disorder in the early life stages, it is only around adolescence and early adulthood that the symptoms of psychosis begin to appear (Eranti *et al.*, 2013). Evidence has shown that the onset of the psychotic phase is preceded by an average prodromal phase of 5 to 6 years in 75% of individuals who will also present subthreshold psychotic symptoms for up to 1 year before being hospitalised with an acute psychotic episode (Häfner, 2003; Sørensen, 2009). The succeeding prodromal phase is signalled by the onset of attenuated psychotic symptoms (APS) which are subthreshold symptoms of psychosis or basic symptoms (BS) alongside a decline in functioning. The duration of the prodromal phase can vary between two and five years (Häfner, 2003). The psychotic phase is marked by the onset of full-blown psychotic symptoms,

characterised by repeated psychotic episodes divided by intermittent periods of remission. The worsening of cognitive functioning occurs in the first 5 years from the onset of psychotic episodes. Finally, the stable phase consists of a reduction of psychotic symptoms followed by heightened negative and cognitive symptoms. Remission periods which can occur to various degrees can be observed at each of these stages and can lead to permanent recovery (Andreasen, 1989).

Early detection of psychosis has been acknowledged as a crucial aspect in this field of research given the societal and individual impact of psychosis as well as the correlation between longer duration of untreated psychosis and poorer treatment response outcome (Loebel *et al.*, 1992; Farooq *et al.*, 2009).

Amongst several challenges, one issue with the conceptualisation of retrospective psychosis consists in the necessity of progression to psychotic stage, which can solely be established following formal diagnosis. For this reason, the clinical high-risk state of psychosis [CHR-P; also called At-risk mental state (ARMS) or ultra-high risk (UHR)] has been established by Yung and McGorry (1996) to define a prospective rather than retrospective definition of the prodromal stage, which opened the door to research into the high-risk stages of psychosis.

1.7 The clinical high-risk state for psychosis.

Given the limited efficacy of improving the course of ScZ once psychosis threshold has been reached (Millan *et al.*, 2016), studies in CHR-P individuals have grown exponentially due to the benefits of investigating individuals at early stages of psychosis before the influence of medication and chronicity impacts as a confound in the examination of the disorder (Sisti and Calkins, 2016). CHR-P participants can be identified as presenting APS as well as other symptoms including negative, cognitive symptoms, neurobiological deficits, and functional decline (Fusar-Poli and Politi, 2008; Fusar-Poli *et al.*, 2012).

The construct of CHR-P falls under two sets of criteria: the ultra high-risk (UHR) criteria and the BS (Fusar-Poli *et al.*, 2013). The former detects psychosis when functioning has begun to deteriorate and therefore at a later prodromal stage, yet, BS can be observed before functional impairment and is therefore detected at an earlier prodromal stage.

1.7.1 Basic Symptoms

BS consists of subtle subthreshold self-reported impairments in a series of domains including attention, memory, perception, motor abilities and emotional regulation (Schultze-Lutter *et al.*, 2016). These symptoms are called "basic" because they consist of the most subtle and earliest perceivable symptoms at the basis of psychosis development (Huber and Gross, 1989). BS are not necessarily observable by others given that the individual experiencing the symptoms has full insight into the psychopathological nature of the experience. BS are assessed using the 'Bonn Scale for the Assessment of Basic Symptoms' (BSABS) (Gross, 1987) or with shorter versions of the Schizophrenia Proneness Instrument, Adult version (SPIA; Schultze-Lutter *et al.*, 2007) or the Schizophrenia Proneness Instrument, Child and Youth version (SPI-CY; Schultze-Lutter *et al.*, 2015) for children and adolescents. The SPIA and SPI-CY hold the advantage over the BSABS of including the examination of symptom frequency based on the previous 3 months. Two subscales of the SPIA can be identified which define BS characteristic and uncharacteristic of psychosis: the Cognitive Disturbances scale (COGDIS) and the Cognitive- Perceptive Basic Symptoms scale (COPER). Evidence has shown that the COGDIS subset can detect a more imminent risk of psychosis development with 25.3% of participants transitioning to psychosis who met the COGDIS criteria at baseline versus a 14.4% of transitioning in those meeting COPER criteria (Schultze-Lutter *et al.*, 2015).

1.7.2 UHR Symptoms

On the other hand, UHR symptoms are meant to capture sub-threshold psychotic symptoms right before the emergence of the psychotic episode. UHR status is defined by the presence of at least one of the following: genetic vulnerability in addition to a marked decline in functioning (Genetic Risk and Deterioration syndrome: GRD); an intermittent psychotic symptom that resolves within one week from onset without any treatment, namely a brief limited intermittent psychotic symptoms (BLIPS); and APS which consists of the most common UHR manifestation (Fusar-Poli *et al.*, 2013). Various measures have been developed for the assessment of UHR status, among these we find the Comprehensive Assessment of At-Risk Mental States (CAARMS; (Yung *et al.*, 2002), the Structured Interview for Prodromal Syndromes (SIPS) and the companion Scale of Prodromal Symptoms (SOPS) (McGlashan *et al.*, 2001) and the Basel Screening Instrument for Psychosis (BSIP) (Riecher-Rössler *et al.*, 2008). Evidence suggests that those meeting the UHR criteria present a high likelihood of developing psychosis. Yet, the transition rate has changed drastically over the years standing

at 50% of developing psychosis after 1 year (Miller *et al.*, 2002) to then report a transition rate of 8% 2 years after meeting UHR criteria (Morrison *et al.*, 2012). The divergent transition rate threatens the validity of the preventative approach (Yung and McGorry, 2007) and could represent a problem in the detection and natural course of the CHR-P group (Fusar-Poli *et al.*, 2012).

1.8 Cognitive deficits in CHR-Ps

Neurocognitive deficits that are commonly observed in chronic ScZ emerge prior to formal diagnosis (Keefe *et al.*, 2006). CHR-Ps present a series of neurocognitive deficits although less pronounced than in later illness stages and including domains such as attention, working memory, processing speed, executive function, verbal fluency, visual memory and verbal memory (Bora and Murray, 2014). These deficits have been shown to be particularly pronounced at baseline in those CHR-Ps who later develop psychosis (Ho *et al.*, 2010). Hence, studies have investigated cognitive impairment as a plausible marker of psychosis vulnerability and transition. Linn and Hemmer (2011) observed that cognitive performance at baseline was predictive of later functional outcome in CHR-Ps. Additionally, processing speed abilities were shown to predict 10 % and 7 % of social and role functioning (Carrión *et al.*, 2011) even without considering positive symptoms. The deficit in cognitive abilities was replicated in CHR-Ps samples recruited from the general population, characterised by deficits in motor speed and executive functioning (Haining, Brunner, *et al.*, 2021). Social cognition was found to be a particularly important predictor, with larger effect seizes than other cognitive measures (Fusar-Poli *et al.*, 2012).

1.9 Prevalence

Evidence collected from self-report questionnaires and interviews have shown that APS are frequently reported by help-seeking individuals (Fusar-Poli *et al.*, 2018). It has recently been shown that individuals recruited from the community report subthreshold symptoms at a significantly higher severity and frequency thus meting CHR-P criteria (McDonald *et al.*, 2019; Haining *et al.*, 2020).The prevalence of individuals presenting subclinical psychotic symptoms is around 5% in the general population as suggested by meta-analytical findings (Van Os, 2009).However, the prevalence rate in children and adolescents (9-18 years old) appears to be about 17% and 7.5% respectively (Kelleher *et al.*, 2012).

1.10 Psychosis experiences in the general population.

Studies have shown that psychosis symptoms can be experienced in the general population, the so called psychosis-like experiences (PLEs) and can be detected through self-reported questionnaires and/or clinician rated measures (Kendler *et al.*, 1996). From a relatively early study that recruited 5,000 community-based individuals through an online survey, it was observed that 24% of these individuals presented PLEs, whereas only 1% of these presented a ScZ diagnosis. Metanalytic findings observed that the presence of PLEs in the general population was found to be around 5% (Van Os, 2009). While PLEs can be transitory in nature and provide no distress or significant functional impairment to the individual, the presence of these symptoms can result in later development of psychosis (Poulton *et al.*, 2000; Van Os, 2009). Additionally, similar risk factors are present in individuals with PLEs and the psychotic population, such that there is a higher prevalence of males with PLEs, ethnic minorities and stressors and life traumas being significant influence factor. Higher prevalence of PLEs is observed in individuals exposed to cannabis and other psychoactive drugs (Van Os, 2009).

1.11 Outcome of CHR-P individuals.

Transition rates of CHR-P individuals are set around 20% after 2 years from baseline assessment, whereas transition rates seem to plateau after 2 years (Fusar-Poli *et al.*, 2012; Fusar-Poli, Cappucciati, *et al.*, 2016). Indeed, psychosis transition risk at a 6-year follow-up and later follow-ups was set at 3.2% for participants referred to secondary mental health services (Fusar-Poli, Cappucciati, *et al.*, 2016). The overall transition rate at a 10 year follow-up was at 35% (Nelson *et al.*, 2013), not much far off the initial 20% after 2 years (Fusar-Poli, Cappucciati, *et al.*, 2016).

Rates of remission are found to be around 50% after a 2 - 3 year follow-up and 70% at 6-7.5 follow-ups (Beck *et al.*, 2019). Yet, there is a consistent number of those meeting CHR-P criteria who will continue to experience sub-threshold symptoms for several years. It has been shown that although remission of positive symptoms tends to be around 2 years, non-specific symptoms of psychosis such as mood or anxiety symptoms tend to linger for much longer periods (De Wit *et al.*, 2014). It is still nowadays unclear on what basis recovery can be established and how subthreshold symptoms can impact on recovery. It is very common for CHR-P participants to present comorbidity with an Axis-I diagnosis, especially anxiety and affective disorders (Michel *et al.*, 2018). The prevalence of comorbidity appears to be equal

across remitted and non-remitted CHR-P participants with rates of 63% and 67% respectively (Beck *et al.*, 2019).

1.12 Outcome of CHR-P not transitioning to psychosis.

Yet, as previously mentioned, transition rates have markedly declined in CHR-Ps with psychosis transition currently set at 20% (Fusar-Poli, Schultze-Lutter, *et al.*, 2016) and have become a crucial issue of debate concerning the utility of UHR diagnosis. Such debate is exacerbated by the limited understanding of outcome of those individuals who do not transition to psychosis (Simon and Umbricht, 2010; Schlosser *et al.*, 2012). Remission rates of CHR-Ps who will not transition to psychosis (CHR-NT) have been estimated by metanalytical findings at 73% of 773 CHR-Ps as estimated at a 2-year follow-up. Yet, 46% CHR-NT fully remitted from subthreshold psychotic symptoms, which consisted of 35% of the overall CHR-P sample. These data suggest that CHR-NT are formed by a heterogeneous group consisting of individuals who will remit from prodromal psychotic symptoms, but that also present a subgroup of individuals who will continue to experience APS (Simon *et al.*, 2011; Fusar-Poli *et al.*, 2013).

In addition, only 49% of CHR-Ps that presented the highest functional impairments later transitioned to psychosis, indicating that a large portion of poor-functioning CHR-Ps were non-converters. This finding indicates that intervention should not be limited to those individuals who will later develop full-blown psychosis (Lin *et al.*, 2011). Most research in CHR-Ps has focused on transition to psychosis as the main outcome based entirely on typical psychotic features such as positive symptoms. Yet, there is a need to identify those individuals who present a risk of poor long-term functional outcome (Fusar-Poli *et al.*, 2013).

1.13 Predictors of outcome in CHR-Ps

As discussed above, the primary interest of research in high-risk and early psychosis has been on the examination of predictability and progression from high risk to psychosis onset to enable early treatment (Fusar-Poli *et al.*, 2012). Nonetheless, as mentioned in the present manuscript, the CHR-P cohort is highly heterogeneous and despite presenting an elevated risk of developing psychosis, most of these individuals will not transition within a 2–3-year follow-up (Yung and McGorry, 2007). As a result, there is a substantial body of research that aims to investigate clinical and neuropsychological predictors that may help predict clinical outcomes in CHR-Ps (McGuire *et al.*, 2015).

1.14 Biomarkers in ScZ

Research aimed at the identification of a biomarker for ScZ has been widely investigated. A biomarker consists of an aspect of the disorder of interest that is examined as an indicator of pathogenic processes or biological response to an intervention (FDA-NIH Biomarker Working Group, 2016). In ScZ there is currently no biomarker available for diagnosis or treatment strategies. A large body of research has observed that a reduction in brain-volume, especially in the hippocampus, is associated with ScZ and may therefore constitute a plausible biomarker. Yet, a recent meta-analysis revealed that across studies on hippocampal volume in CHR-Ps there hippocampal volume reduction was not a significant predictor of psychosis transition suggesting that it may not be a useful biomarker in clinical high-risk individuals (Walter *et al.*, 2016; Hinney *et al.*, 2021). The explanation for the limited development of neuroimaging biomarkers in schizophrenia is three-fold: (1) the heterogeneity of the disorder, as more biologically homogeneous subgroups of patients are required (Cuthbert and Insel, 2013); (2) there is little consensus across studies on how to harmonise the imaging sequences and reduce measurement error (Grzenda and Widge, 2020); (3) such biomarkers should hold the ultimate goal of a positive health impact on the patient, whereas simply measuring biomarker performance does not confirm and is not sufficient for clinical utility (Pletcher and Pignone, 2011). Despite decades of research, it is currently not possible on the basis of clinical screenings to predict whether CHR-Ps may or may not progress to frank illness (Morgan *et al.*, 2021). Therefore, a clinically relevant biomarker is needed for the improvement of early diagnosis and ultimately for prevention of psychosis transition.

1.15 Language in mental health.

Language, considered a window into the mind (Pinker, 2003), is widely recognised to carry essential information in psychiatry, being the medium that enables communication between clinician and patient, which facilitates diagnosis and through which therapeutic treatment is delivered. Analysing the person's verbal behaviour comprises a series of advantages: language abnormalities cannot be hidden, emotional and cognitive expression can be easily observed through language and finally language output can shed light on the neural mechanisms of motoric and acoustic variations which can be examined and compared thanks to the similarities in human vocal anatomy (Low, Bentley and Ghosh, 2020). Language has been defined as a multidimensional system underlying human communication which is strongly linked to cognitive functions (Chomsky, 1995) and is associated to a wide number of physiological, psychological, and cultural domains (Kita, 2003). Language is comprised by several

interrelated components including the domains of semantics (meaning of words), syntactic level (grammatical structure of sentences), coherence (logical flow of meaning) and the lexical domain (word level) (Holmlund *et al.*, 2020). Additionally, speech production involves a wide range of areas including a wide network of brain regions among which are visual, auditory and sensorimotor areas (Denes, 1963). such as the motor and somatosensory cortical areas, cerebellum, basal ganglia, and thalamus, as well as regions that are more specialized for speech and language, including inferior and middle prefrontal cortex and superior and middle temporal cortex.

1.16 Language and psychosis.

Being such a valuable source of information, language consists of a promising alternative as a biomarker for prognosis, detection and diagnosis of psychosis (DeLisi, 2001; Pennebaker, Mehl and Niederhoffer, 2003; Covington *et al.*, 2005; Kuperberg, 2010b). As a biomarker, language has the advantage of being quantitatively reproduced by the clinician without specialist training (Tan *et al.*, 2021).

Abnormalities in communication are widely established in ScZ since the first definition of the disorder (American Psychiatric Association, 2013) when Bleuler defined that ScZ patients presented a disorder of thought rather than an impairment of language itself (Bleuler, 1911). However, the term “schizophrenic language” was later coined by Chaika as a set of impairments at the semantic (meaning) syntax (grammar) and phonological levels (Chaika, 1990; DeLisi, 2001; Covington *et al.*, 2005). Interest in language processing abnormalities has increased recently due to the plausible use of speech to classify diagnosis (Corcoran *et al.*, 2020; De Boer *et al.*, 2021). Language processing is particularly important for the understanding and tracking of psychosis, since the affective, cognitive and socio-functional aspects of language are central for the diagnosis of psychotic disorders (Association, 2013; Organization, 2018). However, the speech disturbances encompass a wide range of vocal behaviour domains including but not limited to disorganised speech, presenting incoherent semantic content, inappropriate word choice (echolalia), unusually flat and sparse prosodic content (de Boer, van Hoogdalem, *et al.*, 2020). Among the linguistic impairments observed in ScZ, phonology seems to be intact, where the phonological rules of speech seem to remain even in the most absurd speeches of psychosis patients. Indeed, Chaika mentioned that the patient’s speech can resemble a familiar language which is however not understood by the listener (Covington *et al.*, 2005).

1.17 Neural substrates of language disturbances in psychosis.

Language processing involves a widespread network of multiple brain regions including the inferior frontal, the superior temporal and the middle temporal gyri as well as the superior temporal sulcus and the inferior parietal lobe of both hemispheres. Cortically, Heschl's gyrus (HG) is involved in language processing, visual word form area in the temporo-occipital region for reading and premotor cortex and supplementary motor area are involved in speech production and articulation (Price, 2010).

Abnormalities in the structure of the language network outlined above have been observed in ScZ, characterised by altered activation in the frontotemporal semantic and phonological processing (Kuperberg, 2007), including Broca's and Wernicke's area (Sans-Sansa *et al.*, 2013). Crucially, aberrant white matter (WM) language tract was observed in ScZ patients (Cavelti *et al.*, 2018) and in CHR-Ps (Thermenos *et al.*, 2013; Li *et al.*, 2019), indicating the presence of neural aberrations before psychosis onset. Genetic predisposition for the development of language aberrations has also been identified. Epigenetic evidence has shown the involvement of FOXP2 gene in the development of language impairments in ScZ (Tolosa *et al.*, 2010; Li *et al.*, 2013). Yet, some studies have observed inconsistent results concerning the gene's polymorphism and ScZ (Kang *et al.*, 2008). An additional gene has shown implications in language impairments in ScZ, namely the dysbindin 1 (DTNBP1) which presented associations with neural correlates of language production (Markov *et al.*, 2009). Yet, genetic findings remain preliminarily.

1.18 Formal Thought Disorder.

Many of the language abnormalities observed in ScZ are joined together under the symptom cluster of FTD, which is primarily a disorder of communication rather than a disorder of language (see *Section 1.3.1*; Kuperberg *et al.*, 2008). Accordingly, disorganised speech in ScZ is characterised by impairments at the pragmatic level rather than at the syntactic or semantic level.

FTD represents a cluster of cognitive, linguistic, and affective disturbances occurring in ScZ patients (Kuperberg, 2010a) and is typically divided into two subdomains: positive thought disorder (PTD) and negative thought disorder (NTD; Peralta and Cuesta, 1999). PTD includes reduced semantic coherence, derailment (speech that results off track) and tangentiality (oblique, irrelevant answer to question) as well as loosening of associations (production of word associations unrelated to the discourse). PTD is found to be correlated to other positive symptoms such as delusion (Docherty *et al.*, 2003). NTD includes reduced verbosity (number

of words) and reduced syntactic and semantic complexity characterised by poverty of speech (i.e. alogia) and tends to occur more often in patients with other non-linguistic negative symptoms. NTD has been shown as a better predictor of clinical and functional outcome compared to positive FTD (Roche *et al.*, 2015).

1.18.1 Characteristics and Prevalence of FTD.

The prevalence of FTD , although highly dependent on the assessment methods used to examine the language disturbances, consist of the highest rate in ScZ (50/80%), schizoaffective disorders (60%), depression (53%) and healthy control with a prevalence rate of 6% (Cavelti *et al.*, 2018). FTD was initially considered a specific symptom of schizophrenia, yet FTD can occur in relatives of patients with ScZ, in CHR-P individuals, affective psychosis, other non-psychotic psychopathologies as well as in healthy controls (Andreasen, 1979; Andreasen and Grove, 1986; Morgan *et al.*, 2017). FTD is multidimensional in nature, unspecific to ScZ and manifests itself on a severity spectrum (Andreasen and Grove, 1986). One issue in defining FTD as a core symptom of ScZ consist in the fact that FTD is absent in 73% of all patients with ScZ (Roche *et al.*, 2015). Indeed, regardless of the methods used, not all patients present thought disorder and the degree of FTD in those patients that present FTD varies in severity and specific expression of thought disorder. Furthermore, FTD presents temporal variability, given that the severity of symptoms fluctuates across time mirroring the symptom fluctuations seen in other ScZ symptoms (Tan and Rossell, 2019).

1.19 Methods for the assessment of FTD.

FTD consists of the only psychotic symptom that can be objectively measured, while delusion and hallucinations can only rely on self-reports.

It is therefore necessary to determine how these speech aberrations change over the course of illness. Tan and Rossell (2019) suggested a holistic approach to the investigation of speech abnormalities in psychosis. Additionally, FTD should be assessed in relation to negative symptoms for a more comprehensive understanding of language disturbances in ScZ.

Several approaches can be used for the assessment of FTD, including clinical rating scales, , manual linguistic analysis and Natural language processing approaches.

1.19.1 Clinical rating scales.

FTD in ScZ patients is commonly assessed using clinical interviews such as the Positive and Negative Syndrome Scale (PANSS) (Kay, Fiszbein and Opler, 1987), whist it is examined in CHR-Ps using the Comprehensive Assessment of At-Risk Mental States (CAARMS) (Yung *et al.*, 2005) which assesses “disorganised speech” through objective and subjective ratings or

using the Structured Interview for Prodromal Syndromes/Scale of Prodromal Symptoms (SIPS/SOPS) (Miller *et al.*, 1999). While the CAARMS mainly examines tangentiality and circumstantiality, the PANSS and SIPS/SOPS assess negative thought disorder including poverty of speech and impoverished "emotional expression" such as blunted vocal affect.

However, clinical rating approaches are limited by their reliance on ordinal scale measures based on the clinician's impression or the subjective insight of the patient and hold fair to moderate inter-rater reliability (0.7) (Olsen and Rosenbaum, 2006).

1.19.2 Manual linguistic analysis.

An alternative approach consists of manual linguistic analysis methods, less affected by temporal aspects (Docherty *et al.*, 2003). A classical method consists of the Story Game, reacted as an ecologically valid assessment of natural speech validated and used across various psychopathologies. The participant's task is to listen to two brief stories, retell the story, answer to open-ended questions and create a new story based on an assigned topic. The Story Game is rated using the Kiddie Formal Thought Disorder Rating Scale (K-FTDS) (Caplan *et al.*, 1989) and measures "illogical thinking", "loose associations", "incoherence" (of syntax used), "poverty of content" and low referential cohesion (i.e pronominal " *John* later referred to as *he*". Lower referential cohesion and poverty of content cohesion were predicative of later psychosis transition (Bearden *et al.*, 2011). Nonetheless, manual linguistic analysis is difficult to implement and is quite laborious, therefore most studies employing these approaches consist of small-scale studies or are performed in clinical settings.

1.19.3 Natural Language processing.

To address the challenges observed in the described methods, natural language processing (NLP) approaches provide a fast, reliable measurement of speech disturbances (Elvevåg *et al.*, 2007; Bedi *et al.*, 2015). NLP methods emerge from the development of computational linguistics and are promising for the application of these methods for clinical use, allowing the understanding of psychosis using semantic, syntactic, lexical and acoustic analysis obtained from natural speech (Holmlund *et al.*, 2019). NLP offers the advantage of capturing speech in ecologically valid settings and can be used to objectively measure the speech parameters that reflect the underlying thought aberrations (Cohen *et al.*, 2020).

Novel discoveries on ScZ language have been made thanks to NLP, such that speech in ScZ patients results discordant/disconnected in dyadic conversations (Kuperberg, 2010a) and that CHR-Ps present reduced semantic density as an index of poverty of speech and in addition

psychosis transition was marked by increased use of words related to voices and sounds (Rezaii, Walker and Wolff, 2019).

A variety of NLP methods have been used in psychosis literature. Latent Semantic Analysis (LSA) is a semantic space model widely used to analytically define how the meaning of words are used within a context (Landauer, Foltz and Laham, 1998), representing words as "vectors" allocated within a "semantic space" – the distance of words based on the sematic (De Boer *et al.*, 2021). Elvevåg *et al.* (2007) was the first to use LSA in ScZ patients and HCs, reporting less semantic similarity and more unusual word associations and less semantic coherence in patients and achieving classification accuracy 86% based on the language features.

However, semantic analysis is limited by the necessary implementation on large corpus of data (Spencer *et al.*, 2021). Therefore, other methods have been created to examine language in ScZ, such as non-semantic speech graphs wich examine language connectedness as an index of speech complexity (Sigman and Cecchi, 2002; Mota *et al.*, 2012; Palaniyappan, 2021). In graph analysis words are treated as nodes and their connections as edges (Mota *et al.*, 2012; Mota, Copelli and Ribeiro, 2017). Speech graph connections were highly reduced in ScZ patients compared to HCs. Using graph-based analysis it was possible to define the specificity of language connectedness in ScZ, that were discriminated against manic patients (Mota *et al.*, 2012). Moreover, language connectedness was associated with negative symptoms in FEP and chronic ScZ patients (Mota and Santos, 2015; Mota, Copelli and Ribeiro, 2017).

Furthermore, reduced language connectedness has been linked to brain dysconnectivity measures during resting state fMRI (Palaniyappan, 2021).

1.19.3.1 NLP approaches in the CHR-P population.

Recently, language disturbances have become a topic of investigation in CHR-P participants (Corcoran *et al.*, 2020). Emerging evidence suggests that semantic and syntactic disturbances can not only distinguish ScZ patients from HCs (Elvevåg *et al.*, 2007), but speech markers can help predict later transition to psychosis in CHR-Ps (Bedi *et al.*, 2015; Corcoran *et al.*, 2018; Rezaii, Walker and Wolff, 2019). A variety of methods for NLP have been used to examine language in psychosis risk, which, often used in combination, have shown more effective and sophisticated language-based assessment of thought disorder (Corcoran *et al.*, 2020). Combining LSA as an index of semantic coherence and Part-of-speech (POS) tagging in open-ended interviews to determine baseline patterns, can predict later psychosis transition and outperforming clinical measures (Bedi *et al.*, 2015). Similarly, Corcoran and colleagues (2018)

employed LSA and POS tagging to previously collected transcripts (Bearden *et al.*, 2011) obtaining high prediction accuracy of psychosis onset (cross-validation accuracy of 0.71; Corcoran *et al.*, 2018). Additional NLP studies investigated negative thought disorder in CHR-Ps by assessing poverty of content defied as *semantic density* and used in combination with analysis of semantic speech content could predict psychosis onset among CHR-P individuals with 90% accuracy (Rezaii, Walker and Wolff, 2019). Moreover, speech graphs have been used to assess language connectedness in CHR-P individuals showing that disconnected speech was associated with later transition to psychosis (Spencer *et al.*, 2021). Given the numerous NLP, Morgan *et al.* (2021)recently aimed to determine the weight of these measures in discriminating among FEP, CHR-P and HC participants and observed that the most diagnostic NLP features included semantic coherence and speech graph. Evidence in CHR-Ps demonstrates that semantic and syntactic aberrations present prognostic value in psychosis.

1.20 Acoustic impairments of speech in ScZ.

In addition to the widely investigated semantic and syntactic impairments, acoustic aspects of speech may represent an important marker for the assessment and symptom tracking of clinical features of the disorder (Cohen *et al.*, 2016; Tahir *et al.*, 2019). Research employing computational acoustic analysis has been conducted for over a century and holds the potential to quantify clinically relevant speech disturbances in an automatic and efficient manner (Ben-Zeev and Atkins, 2017). Acoustic impairments in ScZ have been associated with core negative symptoms including blunted vocal affect and alogia (Andreasen, 1984; Cohen, Kim and Najolia, 2013) and are associated with lower social functioning and increased social withdrawal (Parola *et al.*, 2020).

A different approach consists of the usage of automated acoustic analysis to identify acoustic aspects of speech, which given the objective nature of the methods used may obtain greater reliability, sensitivity and validity. Crucially, computerised acoustic analysis taps into communication difficulties of ScZ that may remain hidden from the analysis of syntactics and semantics (Andreasen and Grove, 1986) by tackling cognitive, affective and socio-functional aspects of speech that are central to the diagnosis of psychosis (Association, 2013; Cohen *et al.*, 2015; World Health Organization, 2018).

Cognitive resources have been identified as a possible mechanism at the basis of acoustic deficits, provided that higher cognitive demand increases negative thought disorder in ScZ

patients (Barch and Berenbaum, 1997; Melinder and Barch, 2003) as well as in other psychiatric conditions (Cohen, Mitchell and Elvevåg, 2014). Tasks with higher cognitive and social demand present larger effect sizes for acoustic impairments in ScZ when compared to their healthy counterparts (Parola *et al.*, 2020), suggesting that cognitive and social functioning resources mark the divide between individuals with psychosis and HCs. Moreover, pauses are particularly impaired within dyadic contexts (Rapcan et al., 2010), hinting to the central role of social functioning and communication disfunction for language impairments in psychosis. Although a clearly defined mechanisms at the basis of atypical voice production in ScZ has not yet been identified (Cannizzaro *et al.*, 2004; Matsumoto *et al.*, 2013; Walther *et al.*, 2015; Konopka and Roberts, 2016), the above evidence hints to the presence of possible cognitive, social and emotional mechanisms at the basis of speech abnormalities, rendering acoustic features of speech a promising marker of psychosis (Corcoran *et al.*, 2020).

1.21 Acoustic features for automated acoustic analysis.

The purpose of the acoustic analysis methods is to extract from the speech signal the frequency, quality, and intensity of sounds produced by the air flowing from the lungs , through the vocal tract to finally reach the vocal tract articulators (tongue, soft palate, lips and other structures; Zhang, 2016). In the analysis of vocal expression, an enormous number of acoustic measures can be extracted and analysed from a speech sample. These acoustic features can be distinguished between prosodic and temporal signatures of speech. Some of the classical prosodic features extracted in automated acoustic analysis include the fundamental frequency (f0), the slowest and non-periodic aspect of speech, which determines pitch; the first and second formant frequencies (F1: first formant; F2: second formant respectively), important for vowel expression; intensity, characterising the volume of the sound signal (Zhang, 2016). A variety of temporal features can be extracted from the speech signal and include the absence, presence of signal (such as pause length and speech duration) as well as the number of such events (including number of pauses, number of syllables, articulation rate, etc.; Wennerstrom, 2001).

1.21.1 Temporal features.

Temporal impairments observed in ScZ include more numerous and longer pauses in ScZ patients compared to healthy controls (Rapcan *et al.*, 2010). Pause duration and percentage of time speaking have been shown diagnostic accuracy of 81.3% in distinguishing ScZ from HCs (Tahir *et al.*, 2019) and correlates robustly with negative symptoms (Alpert, Kotsaftis and

Pouget, 1997; Rapcan *et al.*, 2010; Cohen, Kim and Najolia, 2013). Pause length consists of the most robust acoustic correlate of negative symptoms, particularly symptoms such as alogia and flat affect have been highly associated with pause behaviour (Alpert, Kotsaftis and Pouget, 1997; Cohen, Kim and Najolia, 2013; Stanislawski *et al.*, 2021). Indeed, patients with flat affect speak often with less inflection and less fluently (Alpert *et al.*, 2000). In addition, previous meta-analytic findings of 13 studies revealed pause duration to have the most robust effect size among all acoustic variables included in the studies (Cohen, Mitchell and Elvevåg, 2014). Crucially, studies investigating speech between dyads have shown particularly pronounced impairments for pauses occurring during dyadic turn-taking in ScZ (Alpert, Kotsaftis and Pouget, 1997; Rapcan *et al.*, 2010; Bedi *et al.*, 2015). Yet, turn-taking behaviour is loosely associated with alogia or other negative symptoms, whilst is strongly linked to positive symptoms such as derailment and tangentiality (Alpert, Kotsaftis and Pouget, 1997). Disturbances in thought processes have been investigated as a plausible mechanism at the basis of temporal impairments in psychosis (Çokal *et al.*, 2019). Aberrant pauses may originate from impairments at the level of lexical and concept retrieval (Hartsuiker, Pickering and Veltkamp, 2004). Accordingly, cognitive resources have shown to be central in the emergence of acoustic aberrations, given that tasks with higher cognitive demand may create or emphasise temporal aberrations including pause duration in psychosis compared to HCs (Cohen *et al.*, 2016). The proportion of spoken time, pause duration, pitch variability and speech rate differed significantly between groups.

1.21.2 Prosodic features.

1.21.2.1 Pitch variation.

The most common prosodic feature shown to be deviant in ScZ consists of pitch variation (Parola *et al.*, 2020). Reduced variation in pitch has been shown in psychotic individuals (Bernardini *et al.*, 2016; Cohen *et al.*, 2016; Martínez-Sánchez *et al.*, 2017; Compton *et al.*, 2018) and has been associated with increased severity of negative symptoms, especially flatten vocal affect (Cohen, Kim and Najolia, 2013). Reduced pitch variation has shown high predictability of psychosis, with studies reporting classification accuracy in differentiating ScZ patients from HCs ranging between 86% and 90% (Cohen *et al.*, 2020; De Boer *et al.*, 2021). However, other studies failed to show pitch as a predictive feature in the classification of ScZ patients (Tahir *et al.*, 2019) or in determining diagnosis across different psychopathologies including schizophrenia and affective disorder (Cohen *et al.*, 2019). Additionally, some

evidence failed to observe an association between pitch contour variability and negative symptoms (Rapcan *et al.*, 2010; Covington *et al.*, 2012). The inconsistent findings obtained from the bulk of literature in ScZ have been further confirmed by meta-analytic findings that revealed only a weak association between pitch variability and flat affect (Parola *et al.*, 2020).

1.21.2.2 Formant values.

Other widely evidenced prosodic features include formant values, crucial markers for the identification of pathological voices (Srinivasan, 2018). Formants are frequency peaks emerging from the vibration of the vocal tract (Zhang, 2016) and are especially prominent in the pronunciation of vowels. The F1 relates to the opening and closing of the jaw and mouth, captured by tongue height, whilst the F2 is characterised by the tongue position and lips rounding (Compton *et al.*, 2018).

1.21.2.3 Vowel space.

ScZ patients present reduced vowel space and lower variability in the first two formants has been associated with worse negative symptoms (Covington *et al.*, 2005; Bernardini *et al.*, 2016). Crucially, vowel space is informative of the individual's emotional state, rendering it a diagnostic measure of negative symptoms of psychosis per se (Yildirim *et al.*, 2004). Nonetheless, even in the case of formant values, other studies have not been able to replicate the abnormal formant effects in psychosis (Arevian *et al.*, 2020).

1.21.2.4 Vocal instability.

In addition, pitch (vocal) instability consists of another prosodic feature, which has however received little attention in the ScZ literature until recently (Agurto *et al.*, 2020). Measures of pitch instability include jitter, shimmer and harmonics to noise ratio (HNR). Jitter is characterised by variations in frequency of the repetitions of sound wave defined by the cycles of opening and closing of the glottis (Teixeira, Oliveira and Lopes, 2013). Simmer is defined by the amplitude changes from cycle to cycle (Zwetsch *et al.*, 2006). The HNR reflects the efficiency of speech; it consists of the ratio between the vibration of the vocal cords and glottal noise (Boersma, 1993). Stressor-provoked anxiety is highly associated with pitch perturbation resulting in the so-called jittery voice (Fuller, Horii and Conner, 1992; Mendoza and Carballo, 1998; Cohen *et al.*, 2016). Thus, despite the limited evidence in the psychosis literature, pitch

instability may be a useful index of arousal level and hostility and psychosis fluctuations within an individual, symptoms which are crucial to monitor. Pitch instability may be a useful marker of psychosis, being an indicator of social and emotional functioning, crucial to the disorder (Cohen *et al.*, 2016). Across the few existing studies within the ScZ literature that investigated this feature, vocal instability, differently from other prosodic metrics, has been associated with non-psychotic symptoms such as clinically rated hostility, depression and anxiety (Cohen *et al.*, 2016).

1.22 CHR-Ps.

The presence of language aberrations in ScZ hints to the presence of these speech features in patients with frank illness, however, it holds poor prognostic value. Thanks to the development of computational methods and the emergence of speech features as possible markers of psychosis, research has recently begun to investigate whether acoustic indices can be predictive at earlier stages of psychosis, in order to improve aetiological understanding of the disorder (Corcoran *et al.*, 2020). To date, there is limited and still emerging evidence that examined temporal or prosodic impairments in CHR-Ps. Temporal impairments include increased number and duration of pauses (Stanislawski *et al.*, 2021).

Although between-turn pauses correlated with positive symptoms within the CHR-P cohort, no group effect for between-turn pauses was observed between CHR-P participants and HCs (Sichlinger *et al.*, 2019). A single study examined a broad range of temporal and prosodic features in CHR-Ps and found only prosodic measures to be significant in the prediction of psychosis transition with high accuracy (90%), outperforming classification using clinical variables only (Agurto *et al.*, 2020). Therefore, the presence of vocal metrics being associated with clinical psychotic symptoms as well as being diagnostic of psychosis risk suggests that acoustic features are a promising window into psychosis emergence.

In accordance with the findings in ScZ patients, studies in the CHR-P population have observed associations between acoustic features and clinical symptoms. Temporal (Stanislawski *et al.*, 2021) and prosodic (Agurto *et al.*, 2020) impairments were associated with negative symptoms of psychosis such as blunted vocal affect or alogia, while aberrant pauses in turn-taking were found to correlate with positive symptoms in CHR-Ps (Sichlinger *et al.*, 2019).

Consistent with evidence in ScZ, these findings suggest that acoustic metrics may constitute biomarkers for early detection and diagnosis (Corcoran *et al.*, 2020). Crucially, predictive value of at-risk psychosis has been found to be suboptimal when solely based on clinical symptoms (presenting a conversion rate of 36% within a 3-year follow-up window; Fusar-Poli *et al.*, 2013). Thus, the emergence of evidence that acoustic measures of speech could provide understanding of psychosis risk and progression.

1.23 Methodological implications emerging from the literature.

Nonetheless, the existing evidence of acoustic impairments in CHR-P individuals is still limited, widely diverse and unsystematic in the methods used, as suggested by recent meta-analytic findings (Parola *et al.*, 2020). Such heterogeneity emerges from different sample size and group comparisons, feature extraction and analysis approaches, limiting the comparability across studies (Corcoran *et al.*, 2020) and plausibly explaining the mixed findings observed within the literature (Eyben, 2016).

1.23.1 Unstandardised acoustic analysis method.

There is currently no established standardised method for eliciting speech for acoustic analysis and previous studies in ScZ have elicited speech using a variety of different methods including verbal fluency tasks, picture description tasks, single-word associations and interviews on general life experiences or clinical-based interviews (Elvevåg *et al.*, 2007). This is problematic because such heterogeneity yields inter-rater differences given that different tasks present different levels of cognitive, social and affective demand that considerably affect speech production. Therefore, studies eliciting more natural, free speech, requiring lower cognitive load may avoid measuring confounds such as cognitive deficits rather than language disturbances (Cummings and Čeponienė, 2010).

1.23.2 Variation in speech sample length.

Studies have used different speech elicitation methods with different speech sample length ranging from ten minutes (Sichlinger *et al.*, 2019) to one hour of duration (Agurto *et al.*, 2020), which can influence the emergence acoustic disturbances given that obtaining a shorted speech sample is likely to yield noisier and less reliable results compared to longer samples which will be more representative of the individual's speech patterns . On the other hand, longer

interviews may be more cognitively demanding resulting in greater acoustic differences between psychosis individuals and HCs (Cohen *et al.*, 2016).

1.23.3 Multitude of acoustic measures.

Despite the large amount of evidence existing on vocal abnormalities in ScZ patients, it is still unclear which of the numerous acoustic measures are most informative for distinguishing psychosis patients from their healthy counterparts (Corcoran *et al.*, 2020). Consisting of a rapidly developing field, automated speech analysis methods led to the proliferation of possible acoustic features that can be examined. Different research groups have developed and used their own sets of features, which as shown by meta-analytic findings only a few of these features overlap across studies (Cohen *et al.,* 2014). Even those studies that examined the same acoustic features have used different inferential methods for feature extraction (mean, standard deviation (sd) vs. distribution, percentiles), as well as incongruent choices of ceiling in pitch extraction and different epochs (ranging from small timescales such as extracting syllables, within utterances or across the whole speech sample; Kiss *et al.*, 2012). Thus, there is a need to define whether those acoustic features and the extraction methods employed in previous studies and which were successful in characterising psychosis, can be used to detect psychosis in different subject pools.

1.24 Aims of this thesis

The present study aims at elucidating the combined contribution of temporal and prosodic features given the mixed findings emerging from the literature. Intriguingly, most studies have considered either prosodic (such as pitch and intonation) or temporal variables (such as speech rate, pauses duration and number). However, only a small number of studies investigated the combined contribution of prosodic and temporal features on the same subject pool and the few studies that investigated both reported for the most part impairments only for either temporal (Pinheiro *et al.*, 2017) or prosodic metrics (Cohen *et al.*, 2007; Bernardini *et al.*, 2016) with limited evidence observing disturbances in both acoustic domains (Rapcan *et al.*, 2010; Martínez-Sánchez *et al.*, 2015) and in neither of them (Pinheiro *et al.*, 2016).

In order to investigate both temporal and prosodic features within the same sample, the present study employed temporal features similarly to the metrics used by de Boer and colleagues (2020) in ScZ patients and prosodic features previously examined by Agurto and colleagues (2020) in CHR-P individuals. The current study aims at extending this evidence by assessing both prosodic and temporal variables in CHR-Ps compared to a group that did not meet the CHR-P criteria but presented other non-psychotic psychopathologies (CHR-N) including substance use, anxiety, mood disorder, eating disorders as well as a group of healthy controls.

The comparison with individuals with non-psychotic disorders was chosen to assess the specificity of acoustic features for psychosis or to establish if these speech features are present across a range of psychopathologies. Mixed findings have shown that features such as pitch variability were characteristic of psychosis when compared to autism spectrum disorder and right hemisphere damage (Fusaroli *et al.*, 2017, 2019). On the other hand, the presence of acoustic disturbances including pitch variation and pause duration impairments have been observed across a range of disorders including psychosis but were also observed in anxiety and mood disorders (Cannizzaro *et al.*, 2004). Nonetheless, the bulk of literature investigating language abnormalities in psychosis has mostly examined speech disturbances in a single disorder, while a limited amount of research assessing acoustic disturbances in psychosis has so far investigated the transdiagnostic nature of acoustic aberrations (Cohen *et al.*, 2019, 2020).

Given the limited sample size across the literature which is on average around 30 ScZ patients (De Boer *et al.*, 2021) and for those few studies on individuals with at-risk psychosis the sample

does not exceed 34 (Agurto *et al.*, 2020), the present study aims to replicate the diagnostic potential of previously used acoustic parameters in a larger sample of 50 CHR-P participants.

Moreover, due to the increasing use of more naturalistic, accessible ways of eliciting speech to obtain sufficiently large and ecologically valid speech samples from participants, the speech samples used in the present study were obtained from previously recorded baseline clinical interviews.

Additionally, we address several methodological aspects. Firstly, while clinical interviews elicit naturalistic speech and simultaneously obtain clinical data (Tahir *et al.*, 2019), they potentially lead to variations in speech durations across groups given the different clinical profiles influencing the acoustic effects. Secondly, given the to-date exploratory approach of studies in the field, numerous and somewhat redundant acoustic features are examined in the bulk of literature. However, feeding an abundance of acoustic features extracted from speech into classification models to determine psychosis risk has its drawbacks. Indeed, although it provides a comprehensive outlook of acoustic characteristics, it comes at the cost of overfitting and limiting the interpretability of the data, since it cannot be clear which and to what extent acoustic features can predict or are associated with psychosis (de Boer *et al.,* 2020).

Finally, with the small subject pools present in the field, cross-validation methods are likely to produce biased estimates of model generalisation given that the examined sample will not be representative of the entire population of interest (Low *et al.,* 2020).

Therefore, to address the first methodological concern, we first used linear regression to remove the influence of interview duration from each dependent variable. Then, we used an iterative approach for variable selection to mitigate the plausible collinearity across the numerous acoustic variables. Moreover, given the relatively small subject pool, although larger than most studies in CHR-Ps, we used bootstrapping for model generalisation.

Based on the presence of prosodic and temporal abnormalities in psychosis (de Boer, Voppel, *et al.*, 2020) and CHR-Ps (Agurto et al., 2020), the questions this thesis aims at answering are the following:

(1) Can temporal and prosodic features discriminate CHR-P individuals and HCs?

(2) Are temporal and prosodic metrics specific to psychosis and therefore can they be observed in CHR-Ps in comparison to CHR-N participants?

(3) Can temporal and/or prosodic features be used to determine diagnostic status?

(4) Can correcting for interview duration affect the group effects observed for temporal and prosodic features?

(5) Do clinical and/or functional symptomatology correlate with temporal and/or prosodic variables within the CHR-P group?

It was hypothesised that: (1) CHR-Ps would present impaired temporal and prosodic features compared to HC; (2) acoustic impairments would be specific to psychosis and would therefore differentiate CHR-P participants from CHR-Ns; (3) temporal and/or prosodic feature would be diagnostic of psychosis risk; (4) acoustic data corrected by speech duration would result in different or absent group effects compared to data before correction of interview duration; (5) Given the association of acoustic features with negative symptoms (Stanislawski *et al.*, 2021) and with positive symptoms (Sichlinger *et al.*, 2019) reported by the literature on CHR-P individuals, acoustic measures would be expected to correlate with symptom severity and functioning within the CHR-P group.

2 Methods

The following chapters of this manuscript consist of an extended version of the accepted article in production listed below:

Bianciardi, *et al.* (in press). Investigating temporal and prosodic markers in clinical high-risk for psychosis participants using automated acoustic analysis. *Early Intervention in Psychiatry*.

2.1. The YouR-study.

This study made use of data previously collected as part of the "Youth Mental Health Risk and Resilience (YouR) Study (Uhlhaas *et al.*, 2017), a longitudinal study that aims to detect neurobiological and psychological characteristics and predictors of psychosis risk. The YouR-study was a project funded by the Medical Research Council (MRC) and had been approved by the NHS Research Ethical Committee Glasgow and Greater Clyde.

Data collection for the YouR-study began in 2014 and was still conducting follow-up assessments of participants throughout 2020/2021. The present research made use of some of the measures collected in the YouR-study, including clinical and functional data obtained during baseline screening; while neuroimaging data (MEG and fMRI), as well as clinical and functional follow-up data, were not used. A flowchart depicting the YouR-Study measures utilised in the current study is depicted in *Figure 1*.

2.2. Recruitment and Participants.

The recruitment process employed in the YouR-study made use of an online-screening approach (see http://www.your-study. org.uk) to identify from the general population CHR-P and CHR-N individuals. To this end, participants were identified via email invitations sent to colleges and universities in Edinburgh and Glasgow as well as via posters and flyers (see McDonald et al., 2019). In addition, participants were also recruited through referrals obtained from NHS patient services in NHS Greater Glasgow and Clyde and NHS Lothian and from other mental health services such as student counselling. Participants' informed consent for referrals was conducted in person or online depending on whether individuals were asked to take part in the online screening first or were directly recruited for in-person assessments.

During baseline interviews, 146 CHR-P individuals that were recruited at the University of Glasgow (n = 109; 74.7%) and Edinburgh (n = 37; 25.3%) sites. The recruitment of CHR-P individuals occurred mostly from the general population however, eight participants at

baseline, five at 6-month follow-up and two at 12-month follow-up were recruited through referral.

Those participants who did not meet the CHR-P status consequent to the first baseline assessments were defined as CHR-N participants and were considered in the study as a single group of individuals with other non-psychotic psychopathologies including: affective disorders, substance and/or alcohol abuse, eating disorder, suicidal ideation or intent. Forty-six CHR-N participants were identified from the study. Furthermore, a volunteer database held by the University of Glasgow's Centre for Cognitive Neuroimaging (CCNi) was utilised for the recruitment of 55 healthy controls.

Among the study's inclusion criteria, all participants were required to complete written informed consent prior to the beginning of the study, had to be between 16 – 35 years old, and have normal or corrected to normal visual acuity. Exclusion criteria that applied to all participants consisted of presenting existing medical or neurological disorder or head injury, having metal implants in any body parts (exclusion criteria required in order to be able to participate in the neuroimaging assessments), pregnancy at the time of the assessment or suicidal ideation or intent. Additionally, if HC and CHR-N participants presented an Axis I diagnosis and/or had a first-degree relative with a diagnosis of a psychotic disorder, they had to be excluded from the study.

During the web-based screening, informed consent was provided online. After having obtained informed consent, participants completed the following online questionnaires: 1) the 16-item version of the prodromal questionnaire (PQ-16;(Ising *et al.*, 2012)) extracted from the original 92-item prodromal questionnaire (PQ;(Loewy *et al.*, 2005)) with the aim of examining the presence of psychotic experiences and 2) a scale consisting of only 9 items which were developed to assess perceptual-cognitive anomalies (PCA) and thus, to detect the presence of basic symptoms (see McDonald et al., 2019). To be eligible to take part in the subsequent part of the study and attend the in-person clinical assessments, participants were required to meet the cut-off scores from the web-based questionnaires. This consisted of scoring 6 or more positively endorsed items on the PQ-16 scale and/or a cut-off of 3 or more positively endorsed items obtained from the PCA scale.

2.3. Baseline clinical assessments.

Participants meeting the pre-determined inclusion criteria (PQ: ≥6 items; PCA: ≥3 items) were asked via email to take part in the baseline clinical assessments to establish CHR-P status and to obtain baseline neuropsychological measures.

2.3.1. Demographic information.

Informed consent for the second part of the study was obtained at the start of the first baseline clinical assessment. Participants were informed of their right to withdraw at any time during the course of the study and that withdrawal from the study would not affect their medical care.

Next, demographic information was obtained from each participant at baseline and consisted of age, gender, years of education, citizenship, housing situation, family history of illness, presence of learning difficulties and experienced birth complications. Moreover, social factors such as smoking history, alcohol, substance use/dependence were also examined as additional demographic information. Any incidence of physical or mental health difficulties in the 12 months antecedent to study was recorded for all participants, if any illness was reported, participants were asked whether they were undertaking any drug or psychotherapy treatment and if in their lifetime they experienced being hospitalised for episodes of mental health difficulties. If at any point of the study, participants reported imminent suicidal ideation or intent, they were excluded from the study and were promptly referred to appropriate support/intervention services.

2.3.2. Assessment of CHR-P status.

In-person clinical assessments were conducted by trained research assistants and postgraduate level researchers. Two semi-structured interviews were used to ascertain CHR-P criteria. Firstly, the presence of attenuated psychotic symptoms was examined using the positive scale of the Comprehensive Assessment of At-Risk Mental States (CAARMS; Yung et al., 2005) and secondly the 14-item Schizophrenia Proneness Instrument, Adult version (SPI-A; Schultze-Lutter et al., 2007) was utilised to assess the presence of a range of basic symptoms (BS) including self-experienced perceptual and cognitive aberrations.

The CAARMS aims to detect ARMS criteria which is defined by high, although not inevitable, risk of developing psychosis (McGorry *et al.*, 1995). The CAARMS identifies individuals with

imminent development of an FEP disorder and determines if the young person meets the cluster of symptoms defined by the ultra-high risk for psychosis (UHR) status according to the CAARMS criteria (Yung *et al.*, 2005). Instead of the full CAARMS scale, the positive (or abbreviated) scale of the CAARMS was administered given that the latter is a widely utilised instrument widely to identify clinically young people who meet the UHR criteria (Orygen, 2015). The positive scale of the CAARMS includes the following subscales: unusual thought content (UTC), non-bizarre ideas (NBI), perceptual abnormalities (PA) and disorganised speech (DS). The rating components for each of the subscales include a 0-6 in intensity (also referred to as global rating) and frequency and duration of symptoms. Moreover, the onset and offset of the experience were recorded and the level of distress was scored on a scale from 0 (not at all distressed) to 100 (extremely distressed). To calculate the overall positive symptom severity score, the sum of the global rating scale score (0-6) was multiplied by the frequency scores (0-6) for each of the four subscales and finally summed across the subscales (Morrison *et al.*, 2012).

To meet the UHR criteria and be classified with CHR-P status, participants had to meet criteria from at least one of the following groups:

1) Vulnerability group (also referred to as trait group): having a Schizotypal Personality Disorder or a first-degree relative with a psychotic disorder.
2) APS: characterised by subthreshold (intensity or frequency), attenuated positive psychotic symptoms experienced in the past year.
3) BLIPS: participants presenting frank positive symptoms but with symptoms duration lasting no longer than a week and spontaneously resolved without treatment.

Additionally, to meet UHR criteria participants were required to have experienced 1) a 30% drop in Global Assessment of Functioning (GAF) score from premorbid level, sustained for one month and occurring in the past 12 months or 2) chronically low GAF (score of ≤ 50) for the past 12 months or longer.

Two items from the SPI-A were used to identify if participants met the CHR-P criteria, namely the COGDIS and/or COPER criteria, reported in Table 1, as identified by the SPI-A measure (Schultze-Lutter *et al.*, 2007). To meet the SPI-A criteria, BS had to be present over the past three months. Symptoms were scored based on frequency (0 – 9), a score between 3 and 6 indicated the presence of BS symptoms. A score of moderate (3, several times in a month or

weekly) to extreme (6, daily but not necessarily continuously) was considered as the symptom is present. Although the symptom was present, a score of 7 would indicate that the symptom was always present, while a score of 8-9 indicated insufficient information to rate the symptom as lower. Participants were asked to rate the distress related to the symptom on a scale from 0 to 100.

Table 1 COGDIS COOPER criteria.

COGDIS criteria	**COPER criteria**
At least two of the following BS:	*At least one of the following BS, which started over 12 months ago:*
Inability to divide attention	Thought interference
Thought interference	Thought preservation
Thought pressure	Thought pressure
Thought blockages	Thought blockages
Disturbance of receptive speech	Disturbance of receptive speech
Disturbances of expressive speech	Decreased ability to discriminate between ideas/perception and fantasy/true memories
Unstable ideas of reference	Unstable ideas of reference
Disturbances of abstract thinking	Derealization
Captivation of attention by details of the visual field	Visual perception disturbances
	Acoustic perception disturbances

Excellent inter-rater reliability (IRR) has been reported for the CAARMS (IRR: 0.85; Yung et al., 2005) as well as for the SPI-A (IRR: 0.91) (Fusar-Poli *et al.*, 2015) when assessed by trained clinicians or researchers.

2.3.3. Assessment of functioning.

During the first baseline assessment, the GAF scale, which was extracted from the SIPS (McGlashan et al., 2010), was employed to evaluate the person's psychological social and occupational functioning in the past month. The measure is scored on a scale ranging from 1 (extremely dysfunctional) to 100 (extremely functioning) and is subdivided into 10 equal parts, each comprising impairments over three areas of functioning: psychological, social and role

functioning. Excellent inter-rater reliability has been shown by the GAF (0.89; Startup, Jackson and Bendix, 2002).

In the second visit of the baseline interview, social and occupational functioning were assessed using the Global Functioning: Social (GF: Social) and Global Functioning: Role (GF: Role) scales (Cornblatt *et al.*, 2007), both derived from the GAF scale. Whist the GF: Social scale examines the participant's social interactions/contacts with friends and family, the GF: Role scale assesses the performance and level of independence in occupational/educational setting depending on the individual's age. These two scales were designed to be complementary to each other and therefore both have range on a score range from 1 (extreme dysfunction) to 10 (extreme functioning). The two scales provide the following 3 scores each: lowest level of functioning in the past three months, lowest and highest level of functioning in the past year. A score between 6-8 is expected for CHR-P individuals on both the GF social and role scales (Carrión *et al.*, 2019). These measures, exclude psychiatric symptoms disentangling these specific areas of functioning and have been suggested as more sensitive measures for the assessment of functioning in at-risk psychosis, given that the GAF scale has shown to be problematic in its psychometric properties since it resulted associated with psychiatric symptoms rather than functioning (Bacon, Collins and Plake, 2002). The separation of the functioning domains and the inclusion of well-anchored descriptors are likely to be at the basis of the excellent inter-rater reliability of these two measures was above 0.75 (Cornblatt *et al.*, 2007).

2.3.4 Additional Baseline assessments.

During the second baseline visit, alongside the functional assessments, participants were administered the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), which assesses 17 of the most common neuropsychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM–IV; APA, 1994) with a prevalence of 0.5% or higher in the general population. In the present study, the disorders that participants met during the MINI assessment were subdivided into the following disorders: mood disorders (major depressive episodes, hypomania and mania), anxiety disorders (including phobias, panic attack disorders and generalised anxiety disorders), eating disorders (anorexia nervosa, bulimia nervosa, binge eating disorder), suicidality, alcohol and/or substance dependence/abuse. The mini presents good test-retest reliability (median kappa = 0.78) and excellent inter-rater reliability (median kappa = 0.92).

The third baseline visit was carried out in order to assess the participant's cognitive functioning. The Brief Assessment of Cognition in Schizophrenia (BACS) was administered to participants (Keefe *et al.*, 2008). The BACS includes the following neurocognitive domains: verbal memory (list learning task); working memory (digit sequencing task); motor speed (token motor task); verbal fluency (semantic fluency, letter fluency); attention and processing speed (symbol coding); executive function (tower of London). Three items of the Penn Computerised Neurocognitive Battery were administered to participants as part of the neurocognitive assessment at the third baseline visit. These included the Continuous Performance Task, the Letter N-Back Test; and the Emotion Recognition Task (Moore *et al.*, 2015).

2.4. Language data acquisition.

The present study analysed previously recorded baseline semi-structured interviews of the positive scale of the CAARMS (Yung et al., 2005) assessments collected as part of the YouR-study between November 2014 and June 2019. Collecting speech from clinical interviews holds the benefits of eliciting highly naturalistic speech samples as well as allowing simultaneous comparison of symptomatology and speech data in an objective, reliable way (Hitczenko, Mittal and Goldrick, 2021). From the Your-study database, available audio recordings of the CAARMS interviews were included in this study. These consisted of interview recordings from 73 CHR-Ps, 24 HCs and 30 CHR-Ns. Exclusion criteria included incomplete CAARMS recordings, interviews where participants spoke for a shorter duration than 2 minutes and/or with unexhaustive participant's speech data (almost exclusively Yes/no answers), resulting in the exclusion of 13 CHR-Ps, 14 HCs and 7 CHR-Ns. The final sample included 50 CHR-Ps, 23 CHR-Ns and 17 HCs.

2.5 Speech recording and pre-processing.

A single "hand-worn" cardioid microphone was placed at an average equal distance from the participant and the interviewer recorded both the participant's and interviewer's speech onto a single stereo channel. The speech was digitally extracted using the Olympus software, ODMS 6.4 dictation module, at a sampling rate of 44,100 kHz with 16-bit quantization.

The digitalised recordings were pre-processed with Audacity® (version 2.4.2; Audacity Team, 2021) and manually separated into two different files containing uniquely the participant's and interviewer's speech. Speech segments where both speakers spoke simultaneously were excluded from the analysis in order to obtain speech files containing solely acoustic properties

from the speaker in question. Pauses resulting from a switch between speakers were assigned to the speaker that started speaking following the pause. The audios were removed from excessive noise and reverb using the *Accusonus ERA Bundle 5.0* plug-in available for the Audacity software. This pre-processing step was performed to improve the quality of the audios since the sub-optimal technical quality of the speech material can significantly affect the analysis result (Teixeira, Oliveira and Lopes, 2013). In addition, all audios were normalised using the *Normalise* script available from the Praat Vocal toolkit - a plugin with automated scripts for voice processing as part of the Praat software (Quené, Persoon and de Jong, 2011). The Normalise command scales the audio's amplitude so that the absolute peak becomes 0.99; this method maximises the audibility of the sound files avoiding distortions.

2.6 Feature extraction.

The digitalised recordings were analysed using the Praat software (Quené, Persoon and de Jong, 2011). In the present study, the acoustic features extracted were chosen based on prior literature and can be grouped into temporal features (in accordance with de Boer et al., 2020) and prosodic measures (in accordance with Agurto et al., 2020).

2.6.1 Temporal analysis.

Speech and articulation rates were automatically obtained from the "Praat Script Syllable Nuclei v2" (Quené, Persoon and de Jong, 2011). The *Praat* output automatically provides the following raw measures: presence and absence of sound signal resulting in phonation time and total pause duration, as well as the number of events consisting of the total number of syllables and the total number of pauses. Pauses were considered as silences longer than 200 m since pauses in speech shorter than 200 ms can still be attributed to breathing and articulation during particular syllable pronunciation such as plosives (e.g. /p/ or /t/, which are voiceless consonant phonemes that present a short, naturally-occurring silence in the sound wave) (Rosen, 1992). The raw measures were adjusted for the total duration of the participant's speech sample or to the total interview (including participant and interviewer speech). Such reference measures were employed given that the raw speech variables are highly dependent on the length of the interview. This resulted in the following outcome variables: articulation rate, speech rate, pause rate, average syllable, pause duration, mean length of runs, the ratio of participant and interviewer speech duration. In addition, the percentages of time speaking, articulating and pausing adjusted to the participant interview duration and percentages of time speaking, articulating and pausing adjusted to the total interview time were obtained.

detailed description and relation to ScZ literature for all the speech features included. Among the temporal variables extracted, articulation rate is a variable that purely describes the speed of speech production (i.e. the motoric aspect of communication), while the rest of the language variables have been shown to involve cognitive processes in the production of speech (de Boer, Voppel, *et al.*, 2020).Temporal measures can be subdivided into breakdown fluency and speed fluency. Breakdown fluency measures the planning of language and therefore is indicative of processing speed; breakdown fluency features reflect the duration and number of pauses as well as time and filled with speech and number of such events (Tavakoli and Skehan, 2005). Speed fluency measures are informative of the level of speech production and speaking skills and include articulation rate, speech rate, mean length of runs and average syllable duration. For an overview of the description and calculation used to define the temporal variables, see *Table 2*.

Table 2 *Temporal features*

Variable	**Definition/calculation**	**Measures**
Articulation rate	Syllables / phonation time	*Speed fluency*. Speed in speech production.
Speech rate	Syllables / speech time (including pauses)	*Speed fluency*. Syllables spoken per seconds
Average syllable duration	Phonation time / syllables	*Speed fluency*. Average duration of syllables
Mean length of runs	Number of silent pauses / Number of syllables	*Speed fluency*. Efficiency in speech production
Pause rate	Total number of pauses / speaking time	*Breakdown fluency.* Number of pauses per minute
Average pause duration	Pause time / number of pauses	*Breakdown fluency.* Mean length of pauses
Percentage of time articulating	(Phonation time / participant interview duration) *100	*Breakdown fluency.* Duration of participant's speech

Percentage of time pausing	(Pause time / participant interview time) *100	*Breakdown fluency.* Duration of participant's pauses
Percentage of time articulating (adjusted)	(Phonation time / total interview duration) *100	*Breakdown fluency.* Duration of participant speaking based on the total interview
Percentage of time pausing (adjusted)	(Pause time/total interview time)*100	*Breakdown fluency.* Duration of participant's pauses based on the total interview
Percentage of total time speaking	(Speech time/total interview time)*100	*Breakdown fluency.* Reflects spontaneity or willingness to speak
Participant/Interviewer ratio	Participant phonation time /interviewer phonation time *100	*Breakdown fluency.* Reflects speech fluency or willingness to speak

2.6.2 Prosodic analysis.

Whilst each recording was divided into participant and interviewer audios, only the participant's audio was utilised for prosodic analysis. Using the *Praat Vocal Toolkit* (Corretge, 2012), for each sound file, silences were automatically trimmed using the *"cut pauses*" command, then pitch information was extracted with the "*extract pitch*" command resulting in a Pitch object extracted from the selected sound file. From the obtained *Sound* and *Pitch* objects, a *Voice Report* output was manually extracted using speaker-specific minimum and maximum frequencies to obtain the following prosodic metrics and appropriate statistical descriptors: glottal pulse period (mean and SD), jitter (local absolute and ppq5; five-point Period Perturbation Quotient), shimmer (local dB and apq5; five-point Amplitude Perturbation Quotient), HNR, NHR, percentage of voice breaks and percentage of unvoiced frames (see *Table 3*). Additional pitch analysis was conducted to determine the fundamental frequency *f0,* the slowest and least periodic aspect of the acoustic signal (Teixeira, Oliveira and Lopes, 2013). The present study made use of a previously implemented method that employs a *Praat* script for pitch extraction and an R script for analysis (Lennes *et al.*, 2016). The *Praat* script utilised

the standard auto-correlation-based pitch algorithm [(command Sound: To Pitch(ac)…) which uses a 40-ms Hanning-filtered window] was applied to study intonation of the participant's speech sample with trimmed silences. To estimate vocal pitch, the distance between consecutive analysis frames was set to 0.02 seconds, yielding 50 pitch values per second. In order to avoid anomalies in the speech data, speaker-specific minimum and maximum frequencies were manually determined (in place of default parameters 50-600 Hz, which are often too far apart for most adult speakers) by inspecting the pitch distributions from the voice report output extracted with the Praat Vocal Toolkit (Corretge, 2012). The R script was modified to obtain the following statistical descriptors of pitch distribution: median, skewness, kurtosis, inter-quantile range (IQR), 5th, 25th, 75th and 95th percentiles, in accordance with the statistical measures obtained by Agurto et al. (2020).

In the present manuscript, all pitch values were converted from Hz to semitones (ST; ST are provided relative to the frequency of 100 Hz) due to the non-linear nature of Hz frequency values. Table 2 provides a detailed overview of the prosodic measures included.

Table 3 *Prosodic variables*

Pitch variation		
Variable	**Description/calculation**	**Statistical descriptors**
Pitch	Repetitive opening-closing sequence of vocal cords. Autocorrelation-based pitch algorithm. Pitch in semitones (relative to 100 Hz).	Median, skewness, kurtosis, 5th, 25th, 75th, 95th, percentiles and Interquartile range(IQR).
Glottal pulse period	Variance in voice quality affected by folds of the vocal cords when speaking	Mean, SD
Voice quality		
Variable	**Description/calculation**	**Statistical descriptors**
Jitter	Fluctuations in pitch	Local absolute, ppq5
shimmer	Fluctuations in volume	Local dB value, apq5

Voice Breaks	Measures the maintenance in phonation during speech	Percentage of locally unvoiced frames, percentage of voice breaks
Harmonics to Noise Ratio	Efficiency in speech. Ratio vibration of vocal cords/glottal noise (dB)	HNR, NHR

Note: SD, standard deviation; apq5, five-point Amplitude Perturbation Quotient; ppq5, five-point Period Perturbation Quotient; dB, decibels; HNR, Harmonics to noise ratio; NHR, Noise to harmonics ratio.

2.7 Statistical analysis.

All statistical analyses were performed using R (version 4.0.5) and R Studio with statistical significance set at $p < .05$ (Lakens *et al.*, 2018).

CAARMS severity was calculated by multiplying the frequency of all four domains by the global score and adding these products, while the frequency scores for each SPI-A basic symptom were summed to obtain the total SPI-A severity. Where no symptom was reported for participants, the frequency and distress scores were set to zero.

Descriptive statistics were calculated as mean and standard deviation for continuous values, median and range for ordinal variables and absolute and relative frequencies for categorical variables. Group differences in baseline demographics and clinical measures were analysed using one-way ANOVAs for continuous variables; non-parametric Kruskal-Wallis H tests for ordinal variables and chi-square tests for categorical variables. Appropriate posthoc tests were performed when required. We observed significant differences in the interview duration across the different groups (CHR-P, 31 minutes; HCs, 15 minutes; CHR-N, 26 minutes). Hence the participant's audio length was regressed out for each of the temporal variables, and interview duration was regressed for prosody. We used linear regression to remove the influence of interview duration from each dependent variable (see *Table 1 & 2*) and took the residual of the model as the corrected dependent variable. For the prosodic measures, the participant's audio duration with trimmed pauses was used to regress out each of the prosodic variables, whrereas for temporal measures, the original length of the participant's audio file (including pauses) was used to regress out each of the temporal variables. This process resulted in two sets of data (corrected and uncorrected) separately analysed for both prosodic and temporal datasets to determine how the interview duration affected the results. Visual inspection of the effect of

interview duration was examined from scatterplots obtained for each of the speech variables before and after correction.

Group differences were investigated for each of the temporal and prosodic variables extracted before and after correction of interview duration. Due to the absence of normally distributed data, detected using Q-Q plots (*Appendix A*), the non-parametric Kruskal-Wallis H tests were used for group comparisons which were performed for prosodic and temporal variables both for uncorrected measures and following correction. This was done to observe how correcting for the length of the interview would affect the effects seen for each speech measure across groups. A follow-up analysis was performed using the Dunn, (1964) Kruskal-Wallis test for multiple comparisons when appropriate and p-values adjusted with the Benjamini-Hochberg method. Descriptive statistics are reported using median and range (see *Table 1* to *4*).

Next, to determine the association between the acoustic variables and symptom severity in CHR-Ps the non-parametric Kendall's Tau Coefficient Correlations was conducted for each acoustic variable and the following clinical and functional scores were measured at baseline in CHR-Ps: the total CAARMS and SPI-A severity as well as the four CAARMS positive subscales (Unusual Thought Content, Non-Bizarre Ideas, Perceptual Abnormalities and Disorganised Speech), GAF, GF: Role and GF: Social scores.

Additionally, it was investigated which of the language variables were predictive of a group membership. Due to the primarily exploratory nature of the analyses, prosodic and temporal variables were entered into binary logistic regressions models with backwards exclusion. Twelve binary logistic regressions were calculated for both language datasets (prosody or temporal) and within each dataset for corrected and uncorrected variables and for each group comparison.

Due to the redundancy of the speech variables (i.e. between variables correlation), an iterative approach for variable selection to mitigate collinearity was implemented. Predictors were allowed to correlate up to four pre-established cut-offs (from .4 to .7) and across 3 or 5 predictors. The correlation limit set to 0.7 has been shown an appropriate a-priori multicollinearity cut-off (Tabachnick and Fidell, 1996). Thus, for each one of the six models, this procedure produced eight models including predictors meeting each multicollinearity criterion. These models were compared based on their Akaike information Criterion (AIC) and the model with the lowest AIC value was selected for further analysis.

The final model was then entered into the binary logistic regression with backward exclusion in addition to the covariates of age, years of education and gender, which are plausible confounds and have been previously controlled for (Agurto *et al.*, 2020; de Boer, van

Hoogdalem, *et al.*, 2020). Additional multicollinearity checks allowed only predictors with Variance inflation factor (VIF) below 5 (Ménard, Skachko and Pannekoucke, 2021), any predictor above this cut-off was excluded from the final regression model. The Nagelkerke pseudo R^2 statistics (NR^2) was used to measure the overall variance explained.

It has previously shown that methods such as k-fold cross-validation may produce biased estimates of model generalisation with small sample sizes (Low, Bentley and Ghosh, 2020). For this reason, we used bootstrapping with the boot function from the car package (Davison and Hinkley, 1997). Bootstrapping is a statistical approach used to estimate the uncertainty associated with a given statistical method by building a sampling distribution of a dataset with replacement (Efron and Tibshirani, 1994). From bootstrapping analysis, 750 resamples were obtained and provided 95% confidence intervals (CI), indicating the distribution of model performance scores based on random sampling. To obtain the final performance metric, we determined diagnostic accuracy utilising the area under the receiver operating characteristic (ROC) curve (AUC). The distribution of model performance scores was additionally visualised in histograms reporting values included within a 95% CI range.

Finally, Antidepressant medications (ADMs) can impair phono-articulation through reduction of the salivary production (Stassen, Kuny and Hell, 1998) Indeed, the CHR-N group presented the highest percentage of participants taking anti-depressants (70% of CHR-Ns), whereas only 16% of CHR-Ps were on ADMs, suggesting that ADMs may consist of a plausible covariate of medication status given that no direct manipulation of medication a-priori was conducted. Four post-hoc linear regressions were performed including group status (Diagnosis: CHR-P vs CHR-N), medication (ADMs or no- ADMs) and their interaction as predictors to examine their relationship with the acoustic features significant from group comparisons. Antipsychotic medication status was examined as a categorical variable (participants taking ADMs vs. participants not on ADMs). Given that participants taking multiple medications were all on ADMs, these were included in the analysis as taking ADMs.

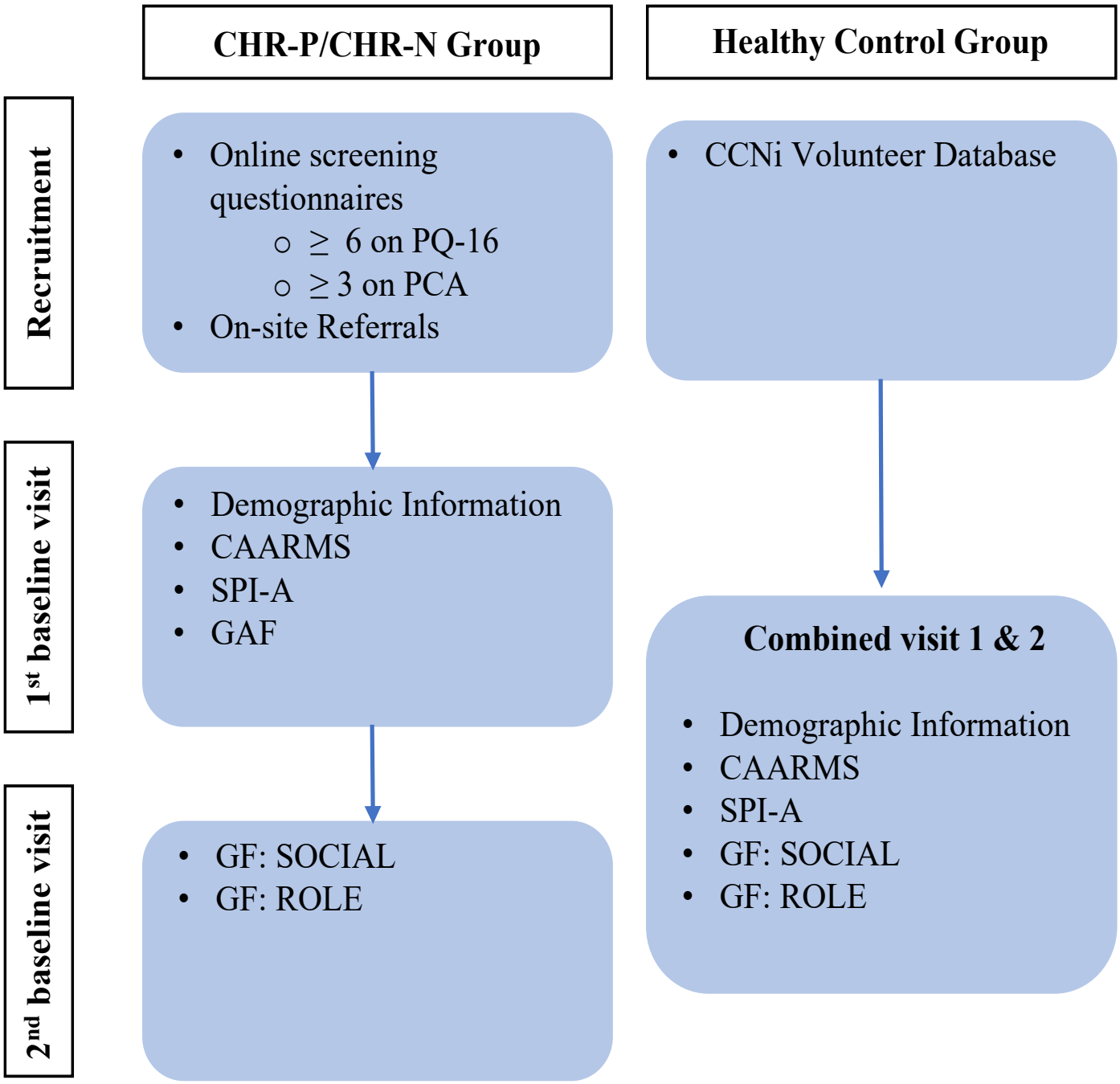

Figure 1 Flow chart of the study protocol

3 Results

3.1 Demographic, clinical and functional information.

Baseline demographic, clinical and functional variables of CHR-P participants, CHR-Ns and HCs are reported in *Table 4*.

Table 4 *Baseline characteristics of CHR-Ps, CHR-Ns and HCs*

Characteristic	**CHR-Ps (n = 50)**	**HCs (n =17)**	**CHR-Ns (n = 23)**	**df**	**F/ χ2/H**	**p**	**Post Hoc Contrasts ***
Age (years), M ± SD	21.1(3.92)	22.4(3.74)	22.5(4.88)	2	F = 1.21	.303	
Gender, N female (%)	39(78.0)	11(64.7)	17(73.9)	2	χ2 = 1.18	.554	
Years of education, M ± SD	14.9(2.74)	16.8(3.72)	16.8(4.26)	2	F = 3.365	.039*	
UK Citizen, N (%)	34(68)	10(58.8)	13(57)	2	χ2 = 1.08	0.58	
GAF, median (range)	61.5(42-95)	89(68-96)	78(43-94)	2	H = 26.3	< .001 ***	1&2,1&3
CAARMS Positive Items, median (range)							
Unusual Thought Content	0(0-5)	0(0)	0(0-3)	2	H = 12.351	< .001 **	1&2
Non-Bizarre Ideas	3.5(0-6)	0(0-2)	0(0-2)	2	H = 25.502	< .001 ***	1&2,1&3

Perceptual Abnormalities	3(0-5)	0(0-3)	0(0-3)	2	H = 29.808	< .001 ***	1&2,1&3
Disorganised Speech	1(0-4)	0(0)	0(0-2)	2	H = 16.947	< .001 ***	1&2,1&3
Total Positive Severity	23.5(0-56)	0(0-12)	4(0-13)	2	H = 44.597	< .001 ***	1&2,1&3
CHR Criteria Subgroup, N (%)							
UHR	18(36)	0(0)	0(0)				
BS	18(36)	0(0)	0(0)				
UHR/BS	13(26)	0(0)	0(0)				
GF: Social, median (range)	8(5-9)	9(8-9)	8(7-9)	2	H = 14.529	< .001 ***	1&2
GF: Role, median (range)	8(4-9)	9(7-9)	8(7-9)	2	H = 10.762	.005 ***	1&2
SPIA severity, median (range)	6(0-37)	0(0-2)	0(0-10)	2	H = 38.617	<. 001 ***	1&2,1&3
Medication, N (%)				6	$\chi 2$ = 16.849	.009 *	

Anti-depressants	8(16)	0(0)	16(69.6)
Other	5(10)	0(0)	0(0)
Multiple	6(12)	0(0)	0(0)
Diagnosis, N (%)			
Anxiety disorders	34(68)	0(0)	13(57)
Mood disorders	30(60)	0(0)	5(21.7)
Eating disorders	4(0.8)	0(0)	1(0.4)
Suicide Risk	24(48)	1(0.6)	4(17.4)
Alcohol Dependence/Abuse	12(24)	0(0)	5(21.7)
Substance Dependence/Abuse	4(0.8)	0(0)	1(0.4)

Note: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; H, Kruskal-Wallis H test; F, F value (ANOVA); $\chi 2$, chi-square test; p, p-value; df, degrees of freedom; N/n, sample size; M, mean; SD, standard deviation.

*1 = CHRs, 2= HCs, 3 = CHR-Ns.

CHR-P participants were characterized by increased symptom severity and distress compared to HCs and to CHR-Ns in their Total positive CAARMS scores and in the following CAARMS subscale items: Non-Bizarre Ideas, Perceptual Abnormalities and Disorganised Speech. For the Unusual Thought Content rating of the CAARMS, CHR-P participants differed significantly only in comparison with HCs, but not with CHR-Ns. Additionally, CHR-Ps had significantly higher symptom severity on the Total SPI-A ratings as well as having significantly poorer global functioning compared to both HCs and CHR-Ns, whist significantly poorer social and role functioning was observed in CHR-Ps compared to HCs only. Group differences were observed for medication status, whereby CHR-Ns reported the highest percentage of individuals undergoing

antidepressant medication (ADM) treatment (69.6%, n = 16 individuals) relative to the CHR-P group (16%, n = 8 participants). CHR-Ns did not report taking any antipsychotic medication nor undergoing any other drug treatment. On the other hand, 5 CHR-P individuals (10%) reported being treated with other medication of various types and 6 CHR-P participants (12%) indicated taking more than one type of medication (which all included ADM in addition to other drugs such as beta-blockers or others). All HCs reported that they were not undergoing any medication treatment. Significant group differences were observed for years of education and medication status, however, the difference in years of education did not remain significant following multiple comparisons. No significant difference was observed among groups for the demographic variables of age, gender and citizenship status.

3.2. Visual inspection of interview duration effect.

Visual inspection of the scatterplots depicting the relationship between each language variable and interview duration before and after correction revealed that most variables were largely robust to the correction. Appendix *Figure 2* depicts those temporal and prosodic variables that from visual inspection presented a partial variation of the acoustic values across groups after interview duration correction. These include the following temporal indices: average pause duration, percentage of time articulating and percentage of time articulating relative to the total interview time (adjusted) as well as speech rate. The change observed in these metrics after interview correction indicated that across the three groups the acoustic values were within a similar range following correction. This was particularly evident for the percentage of time articulating (adjusted), where the values of this temporal metric were within a similar range across groups, whilst before interview duration correction the CHR-P group presented a wider range compared to the other two groups. One additional temporal feature that showed a notable change following interview duration correction consisted of the P/I speech ratio (*panel C; Figure 2*). This feature was characterised by a directional change from a positive to a negative trend, although the relationship between groups remained constant. The negative trend, yet virtually indistinguishable relationship across groups following correction can be explained by the fact that the correction was performed on the speech sample length of the participant but not the interviewer. Although correcting for the interviewer's speech as well would have eliminated the resulting negative trend; such additional analysis was not performed in the present study given that the temporal feature of P/I ratio did not consist of a particularly significant acoustic measure.

Solely one prosodic feature presented some degree of change following correction of interview duration, namely the percentage of unvoiced frames, that following interview duration correction presented virtually indistinguishable values across groups.

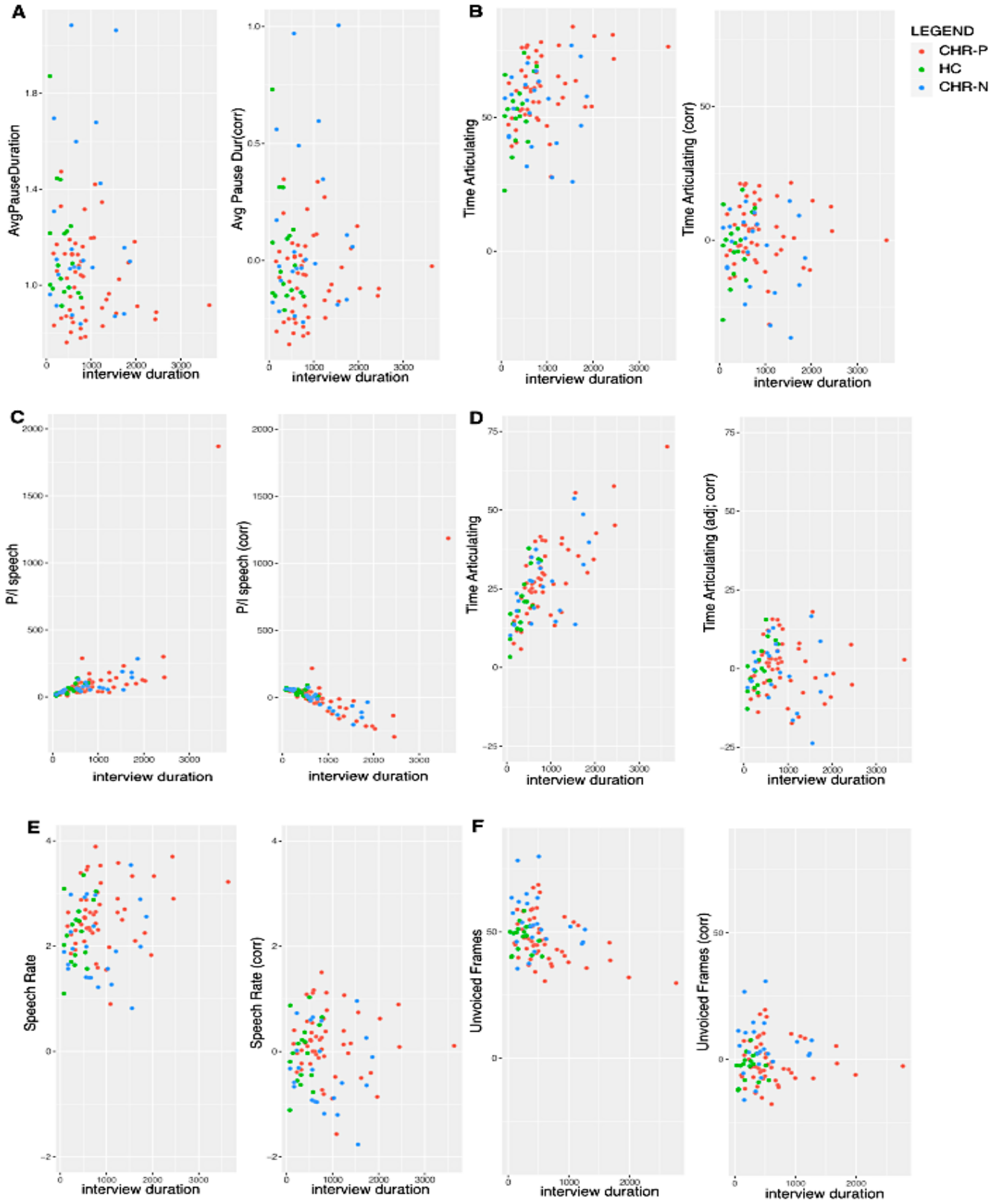

Figure 2 Scatterplots of the relationship between interview duration and acoustic variable before and after correction of the speech sample length.

3.3 Group comparison of acoustic features.

3.3.1. Temporal features.

Table 5 depicts the temporal variables results uncorrected by the interview duration of CHR-P participants, HCs and CHR-Ns.

Table 5 *Group comparison for temporal variables uncorrected by interview duration*

Variable	CHR (n = 50)		HCs (n = 17)		CHR-N (n = 23)				Post Hoc
	Median	Range	Median	Range	Median	Range	H	p	
Speech rate	2.54	(0.90-3.89)	2.28	(1.10-3.35)	1.95	(0.82 - 3.54)	8.86	.012	1&3
Articulation rate	4.28	(3.23 - 5.19)	4.29	(3.68 - 4.91)	3.96	(1.92 - 4.71)	5.03	.081	
Average syllable duration	0.23	(0.19 - 0.31)	0.23	(0.20- 0.27)	0.25	(0.21- 0.52)	5.15	0.07	
Average pause duration	1.027	(0.76-1.47)	1.08	(0.91- 1.87)	1.09	(0.84-2.08)	6.88	0.03	/
Mean length of runs	0.15	(.056-0.56)	0.201	(0.07- 7 0.37)	0.23	(0.075-0.43)	9.34	0.009	1&3
Pause Rate	0.64	(0.22-1.82)	0.81	(0.35 - 1.82)	0.77	(0.34-1.56)	5.31	.070	

Percentage of time articulating	61.16	(27.79-83.92)	53.06	(22.66-74.18)	56.99	(25.93-76.91)	6.40	.041	/
Percentage of time pausing	38.84	(16.08 - 72.20)	46.94	(25.81-77.33)	43.01	(23.09-74.06)	6.40	.041	/
Percentage of time articulating (adjusted)	26.72	(5.84 - 70.18)	19.58	(3.37 - 37.84)	25.61	(10.18 - 53.69)	5.66	0.059	
Percentage of time pausing (adjusted)	16.29	(7.07-39.89)	16.73	(7.07-39.89)	18.43	(7.63 - 58.59)	4.01	0.134	
Percentage of total time speaking	47.004	(12.91-91.79)	37.91	(14.87-68.44)	44.87	(17.80-96.09)	4.59	0.10	
Participant/ Interviewer Percentage	74.92	(9.14-1869.38)	45.44	(9.0-140.69)	57.15	(17.89-284.52)	5.93	0.051	

Note: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; n, sample size; H, Kruskal-Wallis H test. * CHR-P = 1, HC =2, CHR-N = 3

Results before correction of interview duration revealed that CHR-Ps had significantly higher speech rate (M = 2.54) and lower mean length of runs (M = 0.15) compared to CHR-Ns (speech rate M = 1.95; mean length of runs M = 0.23), unexpectedly indicating better temporal performance in CHR-Ps than in CHR-Ns (respectively: $H(2) = 8.86$, $p = .012$; $H(2) = 9.34$ $p = .009$). Group differences were also observed for average pause duration ($H(2) = 8.86$, $p = .03$) such that CHR-Ps had lower values (M = 1.027) compared to HCs (M = 1.08) and CHR-Ns (M = 1.09), percentage of time articulating ($H(2) = 6.40$, $p = .041$) which was higher in CHR-Ps (M = 61.16) , followed by CHR-Ns (M = 56.99) and HCs (M = 53.06) and for percentage of time pausing ($H(2) = 6.40$, $p = .041$). However, the effects seen for average pause duration, percentage of time articulating and percentage of time pausing did not survive multiple comparison correction.

The remaining temporal uncorrected features presented some difference in the median across groups which however did not reach significance and were therefore not reported in this paper.

Group differences for temporal features corrected by the interview duration of CHR-P participants, HCs and CHR-Ns are summarised in *Table 6.*

Table 6 Group comparison for temporal variables corrected by interview duration

Variable	**CHR-P (n = 50)**		**HC (n = 17)**		**CHR-N (n = 23)**				
	Median	**Range**	**Median**	**Range**	**Median**	**Range**	**H**	**p**	**Post Hoc**
Speech rate	0.14	(1.57-1.50)	-0.03	(1.11-1.03)	-0.35	(-1.76-0.96)	7.87	0.019	1&3
Articulation rate	0.21	(-0.92-1.023)	0.11	(-0.52-0.71)	-0.21	(-2.24-0.58)	4.86	0.088	
Average syllable duration	-0.018	(-0.053-0.062)	-0.012	(0.04-0.028)	0.007	(-0.036-0.27)	4.64	0.098	

Average pause duration	-0.068	(-0.36-0.34)	-0.049	(-0.21-0.73)	-0.013	(-0.26-1.00)	4.63	0.099	
Mean length of runs	-0.026	(-0.13-0.39)	-0.004	(-0.12-0.17)	0.029	(-0.10 - 0.27)	7.65	0.022	1&3
Pause Rate	-0.062	(-0.51-1.13)	0.011	(-0.43-0.97)	-0.05	(-0.39 - 0.87)	1.69	0.43	
Percentage of time articulating	3.39	(-31.51-21.44)	-1.03	(29.82-18.83)	-0.47	(-36.51-14.62)	3.133	0.209	
Percentage of time pausing	-3.39	(-21.45-31.50)	1.03	(-18.84-29.81)	0.48	(-14.63-36.50)	3.133	0.209	
Percentage of time articulating (adjusted)	1.22	(-17.35-18.05)	-0.78	(-12.79-15.57)	-1.24	(-23.71-16.64)	0.056	0.97	
Percentage of time pausing (adjusted)	-2.27	(-13.08 - 20.02)	-1.86	(-7.43 - 17.32)	1.25	(-8.59 - 40.20)	6.07	0.048	1&3
Percentage of total time speaking	-2.67	(-23.89 - 23.59)	0.018	(-17.49 - 27.63)	1.91	(-14.69 - 53.12)	2.05	0.360	

Participant/ Interviewer Ratio	-9.04	(-294.04-1186.92)	49.29	(-9.04 - 90.37)	-1.05	(-203.41-74.94)	12.53	0.0019	1&2, 2&3

Note: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; n, sample size; H, Kruskal-Wallis H test. * CHR-P = 1, HC =2, CHR-N = 3

Similarly, to the uncorrected group comparison results for the temporal features, CHR-Ps presented significantly different speech rates ($H(2) = 7.87$, $p = .019$) and mean length of runs ($H(2) = 7.65$, $p = .022$) compared to CHR-Ns, characterised by higher speech rates (*median* =0.14) and lower mean length of runs (median = -0.026) in CHR-Ps (M = -0.026) compared to CHR-Ns (0.029). Additionally, CHR-Ps (M = -2.27) had a significantly lower percentage of time pausing (adjusted to the total interview duration) compared to CHR-Ns (M = 1.25; $H(2) = 6.07$, $p = .048$) as well as lower participant/interviewer speech duration ratio ($H(2) = 12.53$, $p = .0019$), the latter variable was significantly lower in CHR-Ps (M = -9.04) compared to CHR-Ns (M=-1.05) and HCs (M = 49.29). The remaining corrected temporal features presented some difference in the median values across groups which however did not reach significance and were therefore not discussed in the present manuscript.

3.3.2. Prosodic features.

Table 7 displays differences in prosodic features uncorrected by interview duration for CHR-P individuals, HCs and CHR-Ns.

Table 7 *Group comparisons for prosodic variables uncorrected by interview duration*

Variable	**CHR (n = 50)**		**HCs (n = 17)**		**CHR-N (n = 23)**				**Post Hoc**
	Median	**Range**	**Median**	**Range**	**Median**	**Range**	**F**	**p**	
Mean Pulses	0.005	(0.004-0.009)	0.005	(0.003-0.009)	0.005	(0.004-0.009)	2.61	.272	

Sd Pulses	0.001	(0.0005-0.004)	0.001	(0.0005-0.003)	0.001	(0.0004-0.009)	2.10	0.34	
Jitter	0.0001	(0.0005 - 0.0003)	0.0001	(0.00006-0.0003)	0.0001	(0.0007-0.0003)	0.84	.655	
Jitter ppq5	1.24	(0.6-2.83)	1.37	(0.9-3.09)	1.15	(0.77-2.07)	0.84	0.647	
Shimmer	1.29	(0.89-1.79)	1.28	(0.99-1.78)	1.25	(0.99-1.47)	0.433	0.805	
Shimmer apq5	8.30	(4.72-13.89)	8.55	(5.88-13.82)	8.56	(6.02-10.64)	0.096	0.953	
Voice breaks	51.82	(34.52-74.45)	54.90	(43.47-65.37)	58.12	(40.09-82.3)	5.25	0.07	
Unvoiced frames (%)	45.67	(29.76 - 68.62)	49.47	(40.04-58.33)	52.19	(35.45-79.84)	7.29	.003	1&3
NHR	0.19	(0.1-8 0.42)	0.21	(0.13- 0.43)	0.19	(0.11-0.36)	1.797	0.407	
HNR	10.89	(5.17-15.8)	10.16	(5.1- 14.0)	10.36	(6.29-14.72)	0.813	0.666	
Pitch Median ST	11.81	(7.38-25.59)	11.98	(7.63-22.85)	11.43	(5.88-14.47)	2.56	.278	

Pitch Skewness	1.68	(-0.51-3.96)	1.56	(-0.61-4.08)	2.14	(0.24-3.28)	3.53	.171
Pitch Kurtosis	7.39	(1.27-21.53)	5.36	(1.36-24.13)	9.23	(1.45-30.07)	3.14	.208
Pitch 5th pct	-2.13	(-23.82--0.63	-2.65	(-23.06--1.04)	-1.1	(-6.9- -0.5)	2.98	.225
Pitch 25th pct	-0.37	(-20.97-0.64)	-0.7	(-19.27-0.69)	-0.59	(-5.1-1.4)	2.08	.353
Pitch 75th pct	2.72	(-0.62-20.07)	2.6	(-0.33-22.41)	2.43	(0.58-20.39)	1.03	.596
Pitch 95th pct	9.49	(1.74-23.55)	14.0	(2.24-24.69)	10.83	(2.84-24.00)	0.18	.913
Pitch IQR	3.39	(1.54-20.35)	3.55	(2.16-21.71)	3.22	(1.21-19.09)	1.72	.424

Note: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; n, sample size; H, Kruskal-Wallis H test; IQR, Interquartile Range; pct, percentile; ST, semitones; NHR, noise to harmonics ratio; HNR, harmonics to noise ratio; apq5, ; ppq5, ;SD, standard deviation.

* CHR-P = 1, controls =2, CHR-N = 3

Group comparison for prosodic measures uncorrected for interview duration revealed that solely one prosodic variable was significantly different across groups, namely unvoiced frame percentage ($H(2) = 7.29$, $p = .003$) significantly higher for CHR-Ns ($M = 52.19$) compared to CHR-Ps ($M = 45.67$). The group comparisons for the remaining uncorrected prosodic features were not reported given that despite presenting some difference in the median values across groups these prosodic features did not reach significance.

Group differences for prosodic features corrected by the interview duration of CHR-P participants, HCs and CHR-Ns are summarised in *Table 8.*

Table 8 *Group comparisons for prosodic variables corrected by interview duration*

Variable	CHR-P (n = 50)		HC (n = 17)		CHR-N (n = 23)				
	Median	**Range**	**Median**	**Range**	**Median**	**Range**	**H**	**p-value**	**Post Hoc**
Mean Pulses	-0.0005	(-0.002-0.003)	-0.0006	(-0.002-0.004)	-0.0003	(-0.001-0.003)	2.66	.263	
SD Pulses	-0.0002	(-0.0008-0.003)	-0.0001	(-0.0009-0.002)	-0.0003	(-0.0009-0.001)	1.93	.382	
Jitter	-0.00001	(-0.00008-0.0001)	-0.00001	(-0.00007-0.0002)	-0.00001	(-0.00006-0.0001)	0.16	.923	
Jitter ppq5	-0.055	(-0.79-1.49)	-0.053	(-0.53-1.71)	-0.13	(-0.58-0.64)	0.78	.676	
Shimmer	0.004	(-0.42-0.49)	-0.030	(-0.31-0.478)	-0.047	(-0.29-0.17)	0.48	.787	
Shimmer apq5	-0.084	(-3.87-5.38)	-0.066	(-2.74-5.27)	-0.030	(-2.36-2.09)	0.12	.94	
Voice breaks	0.40	(-5.32-5.37)	-0.24	(5.38-3.61)	-0.25	(-4.17-4.24)	5.58	.062	
Unvoiced frames	-3.09	(-17.81-19.5)	-3.10	(-12.29-7.53)	3.004	(-16.154 30.81)	8.89	.001	1&3, 2&3

NHR	-0.02	(-0.12-0.20)	-0.006	(-0.09- 0.214)	-0.017	(-0.105-0.14)	1.12	.570
HNR	0.395	(-5.3-5.37)	-0.248	(-5.38-3.61)	-0.249	(-4.167-4.24)	0.69	.709
Pitch Median ST	-0.31	(-4.62-13.32)	-0.33	(-5.003-10.27)	-0.33	(-6.62-2.15)	1.14	.566
Pitch Skewness	0.012	(-2.16-2.32)	0.0002	(-2.12-2.59)	0.365	(-1.39-1.55)	2.48	.288
Pitch Kurtosis	-1.30	(-7.73-13.42)	-2.36	(-6.7-16.65)	0.43	(-6.61-21.6)	2.2	.33
Pitch 5th pct	1.517	(-19.91-2.90)	1.25	(-19.27-3.06)	1.74	(-3.117-3.66)	1.82	.402
Pitch 25th pct	1.143	(-19.03-2.21)	1.21	(-17.42-2.78)	1.21	(-3.26-3.33)	0.71	.700
Pitch 75th pct	-1.54	(-5.15-15.61)	-2.004	(-4.83-17.82)	-2.030	(-4.008-15.87)	1.91	.385
Pitch 95th pct	-1.41	(-9.95-12.20)	2.01	(-9.62-12.86)	-0.73	(-8.57-12.33)	0.07	.965
Pitch IQR	-2.42	(-4.84-13.87)	-2.91	(-4.79-15.04)	-2.91	(-4.86-12.65)	2.03	.361

Note: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; n, sample size; H, Kruskal-Wallis H test; IQR, Interquartile Range; pct, percentile; ST, semitones; NHR, noise to harmonics ratio; HNR, harmonics to noise ratio; apq5, ; ppq5, ;SD, standard deviation. * CHR-P = 1, HC =2, CHR-N = 3

In line with the uncorrected prosodic findings, group comparison for prosodic measures after correction of variables for the interview duration yielded significant difference across groups only for unvoiced frame percentage ($H(2) = 8.89$, $p = .001$) which was higher for CHR-Ns (median = 3.004) compared to CHR-Ps (M = - 3.09) and HCs (M = -3.10).

The remaining prosodic features that did not reach significance obtained some median difference across groups which however was not strong enough to reach significance.

3.4. Regression analysis.

Binary logistic regressions with backwards exclusion were fitted to investigate whether the temporal and/or prosodic features before and after correcting for interview duration could predict diagnostic status. The fitted models included the following combinations: temporal or prosodic features, corrected or uncorrected data and these four possible combinations were tested for each group comparison (CHR-P vs. HC / CHR-P vs. CHR-N / CHR-N vs HC) resulting in a total of twelve binary logistic regressions. All models initially included the combination of temporal or prosodic variables, corrected or uncorrected data, group comparison as well as the covariates of gender, age and education level.

Table 9 summarises the regression models for temporal variables uncorrected by interview length for prediction of group status.

Table 9 *Binary Logistic Regression models from temporal uncorrected variables for prediction of Group Status*

Model for prediction of CHR-P vs. CHR-N status					
Variable	***B***	**SE**	**Wald**	**p-value**	**OR (95% CI)**
intercept	8.96	2.50		<.001***	0.008 (0.92-6e+6)
Average Syllable Duration	-13.05	6.54	3.99	.046*	2.14e-06(1e-12-1.7e-01)
Average Pause Duration	-2.69	1.19	5.10	.024*	0.07(0.005-0.6)
Years of Education	-0.12	0.09	2.06	.15	0.89(0.74-1.04)
Model for prediction of CHR-P vs. HC status					
intercept	2.28	1.68		0.174	9.786(0.39-303.04)
I/P speech	0.02	0.009	4.08	.044*	1.018(1.009-1.038)
Gender	0.43	0.67	0.41	.519	1.54(0.39-5.65)

Years of Education	-0.17	0.096	3.28	.070	0.84(0.68-1.007)
Model for prediction of CHR-N vs. HC status					
intercept	-9.31	4.03		0.021*	0.0001(7.6e-09- 0.078)
Average Syllable Duration	29.03	14.67	3.915	0.048*	4e+12(0.0034 - 1.8e+27)
Tot Interview Time	0.059	0.029	4.181	0.041*	1.06(1.01- 1.13)

Note: CHR-P, clinical high-risk; CHR-N, clinical low risk; HC, healthy controls; *B,* beta coefficient; SE, standard error; OR, odds ratio; 95% CI, 95% confidence interval. p < .05*; p < .01** ; p< .001***

The model predicting group status for CHR-Ps versus CHR-N accounted for 28% of the variance (NR^2 = 0.281) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.610). The final model presented no source of multicollinearity among predictors (VIF: 1.002-1.008). This model included two significant negative predictors, average syllable duration (β = -13.05, OR = 2.14e-06, 95% CI = 1e-12-1.7e-01, p = .038) and average pause duration (β = -2.78, OR = 0.07, 95% CI = 0.005-0.6, p = .022), indicating greater likelihood of falling within the CHR-P category for those individuals presenting lower average pause duration and lower syllable duration. Additionally, the model included the non-significant predictor of years of education (β = -0.12, OR = 0.89, 95% CI = 0.74,1.04, p = .15). *Figure 3* depicts the receiver-operating characteristic (ROC) curve for the model predicting group status for CHR-Ps versus CHR-N. The area under the curve for the model was 0.774 indicating acceptable discriminative ability of the model.

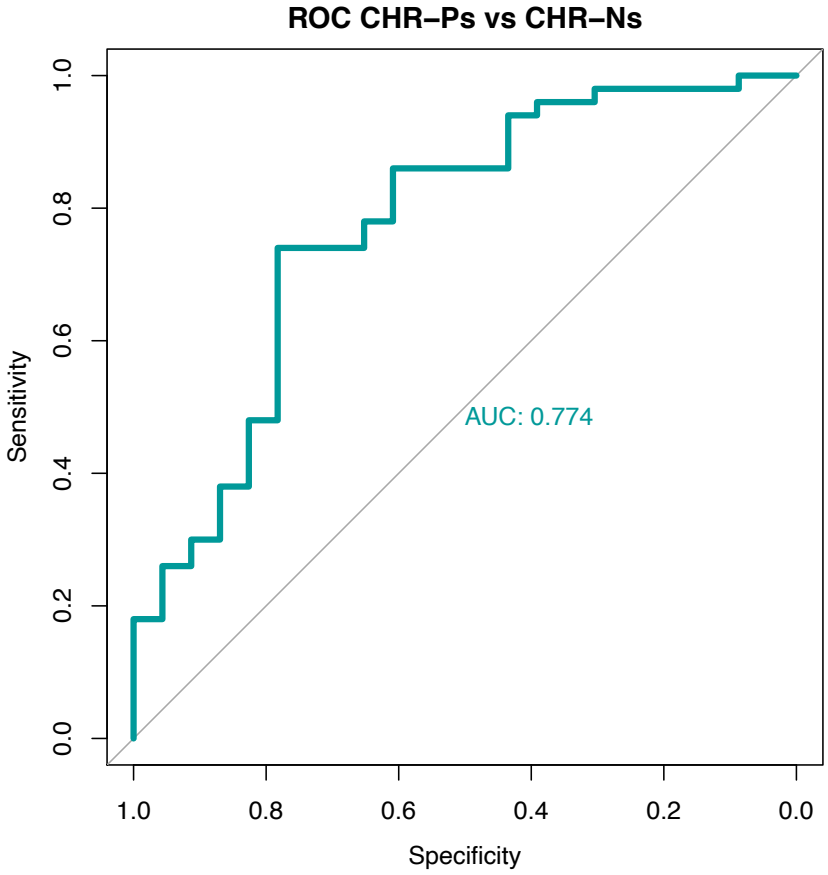

Figure 3 ROC curve for the logistic regression model for temporal uncorrected data discriminating CHR-Ps vs CHR-Ns.

Yet, bootstrapping analysis showed modest model accuracy with Area under the curve (AUC) presenting values not different from chance performance (AUC 95% CI = 0.45-0.67). Whilst average pause duration (95% CI = -5.19, -0.49) remained significant, average syllable duration (95% CI = -33.06, 1.2) and years of education resulted not-significant predictors (95% CI = -0.35, 0.11) following bootstrapping (*Figure 3*).

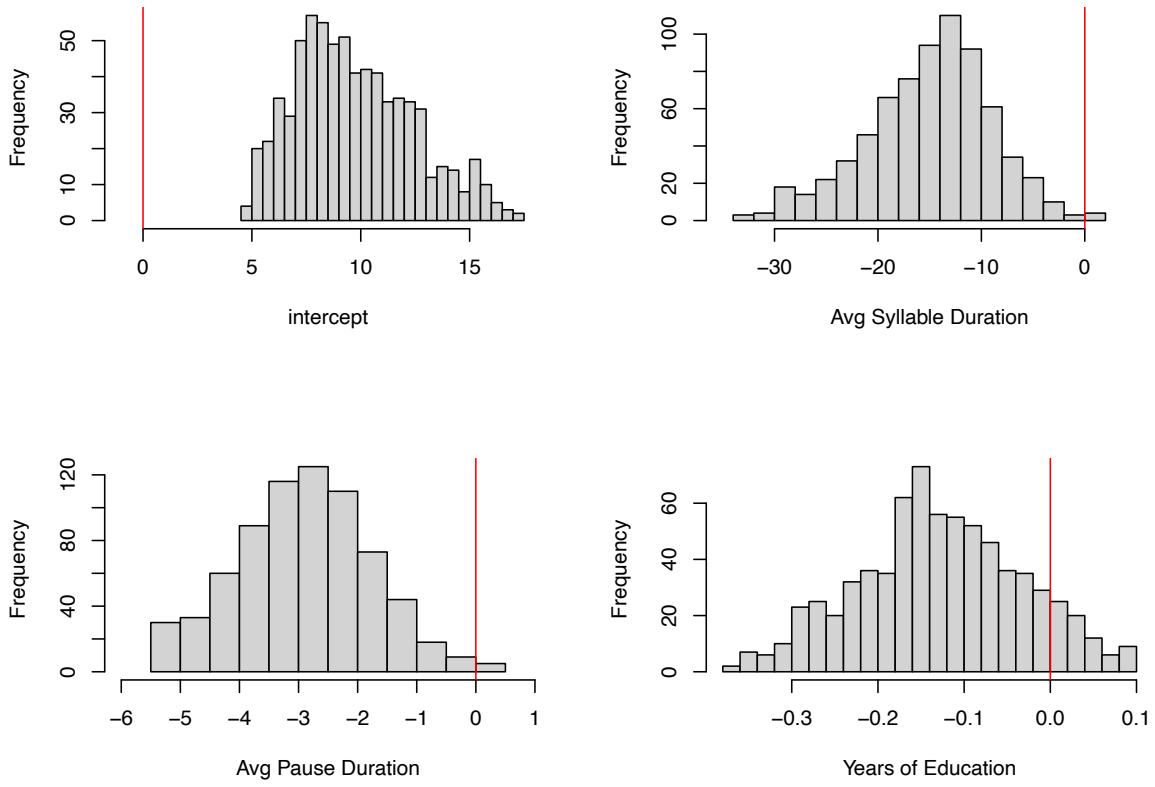

Figure 4 Histograms of the bootstrapping results for the binary logistic regression model of temporal uncorrected acoustic variables determining CHR-Ps and CHR-Ns. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The model determining CHR-N versus HC status accounted for 35% of the variance (Nagelkerke approximation: R2N = 0.350) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.61). Predictors included average syllable duration (β = 29.03, OR = 4e+12, 95% CI = 0.0034-1.8e+27, p = .048) and percentage of participant's speech duration relative to the whole interview (Total interview duration; β = 0.059, OR = 1.06, 95% CI = 1.01- 1.13, p = .041) which were both significant predictors of the model. Predictors presented low VIF values indicating no multicollinearity among predictors (VIF: 1.084). *Figure 4* shows the ROC curve for the model predicting group status for CHR-Ns versus HCs. The discriminative ability of the model was acceptable, with an overall accuracy of 0.79 (AUC).

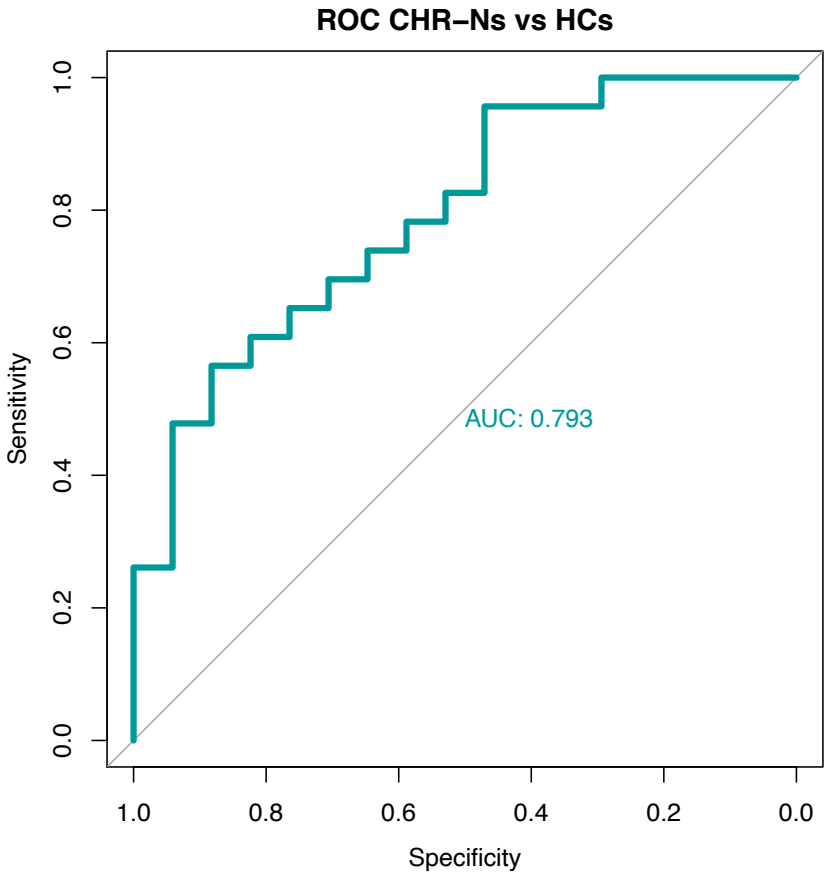

Figure 5 ROC curve for the logistic regression model for temporal uncorrected data discriminating CHR-Ns vs HCs.

Following bootstrapping, only syllable duration remained a significant predictor (95% CI = 10-81.66), while interview duration revealed not significant (95% CI = -0.01-0.17). Yet, the overall diagnostic accuracy included values below chance level after bootstrapping analysis (AUC 95% CI = 0.45-0.70). (*Figure 5*).

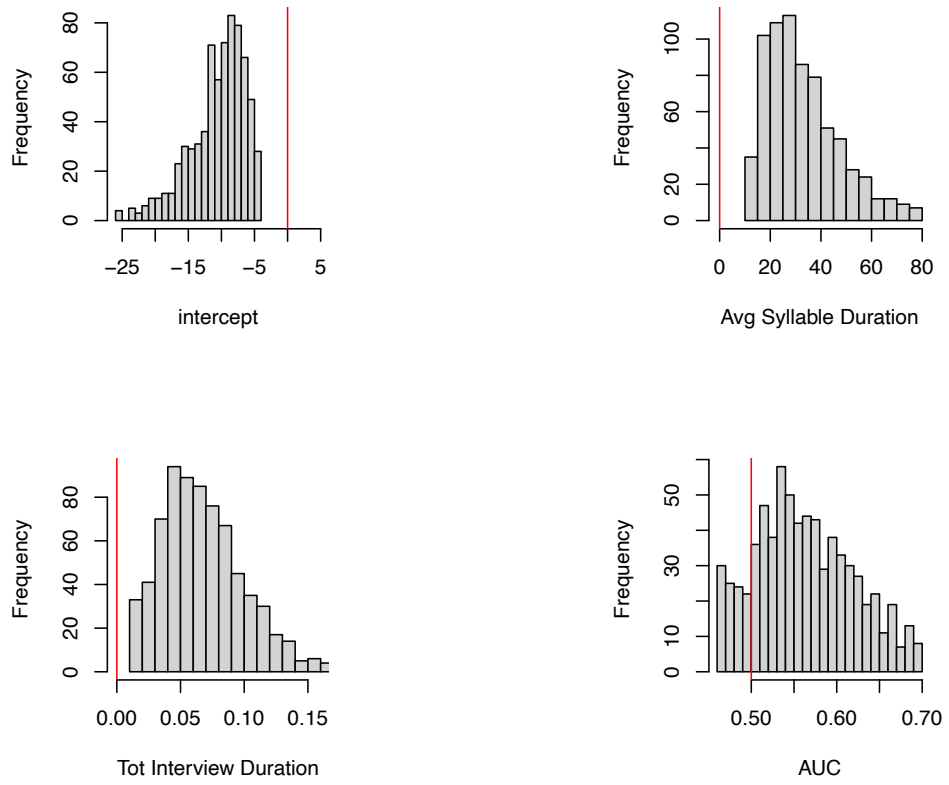

Figure 6 Histograms of the bootstrapping results for the binary logistic regression model of temporal uncorrected acoustic variables determining CHR-Ns and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

In addition, the temporal uncorrected model determining CHR-Ps versus HCs diagnostic accuracy, presented a worse fit compared to the other models, as it accounted for 22% of the variance (NR^2 = 0.220) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.59). This model included participant/interviewer speech ratio (P/I speech) as a significant predictor (β = 0.02, OR = 1.018, 95% CI = 1.009 - 1.038, p = .044), while gender (β = 0.43, OR = 1.54, 95% CI = 0.39 - 5.65, p = .519) and years of education (β = -0.17, OR = 0.84, 95% CI = 0.68 - 1.007, p = .07) were non-significant predictors of the model. No source of multicollinearity was detected among the final predictors (VIF: 1.002-1.016).

The ROC curve for the temporal model predicting group status for CHR-Ps versus HCs is shown in *Figure 6.* The discriminative ability of the model was acceptable, with an overall accuracy of 0.76 (AUC).

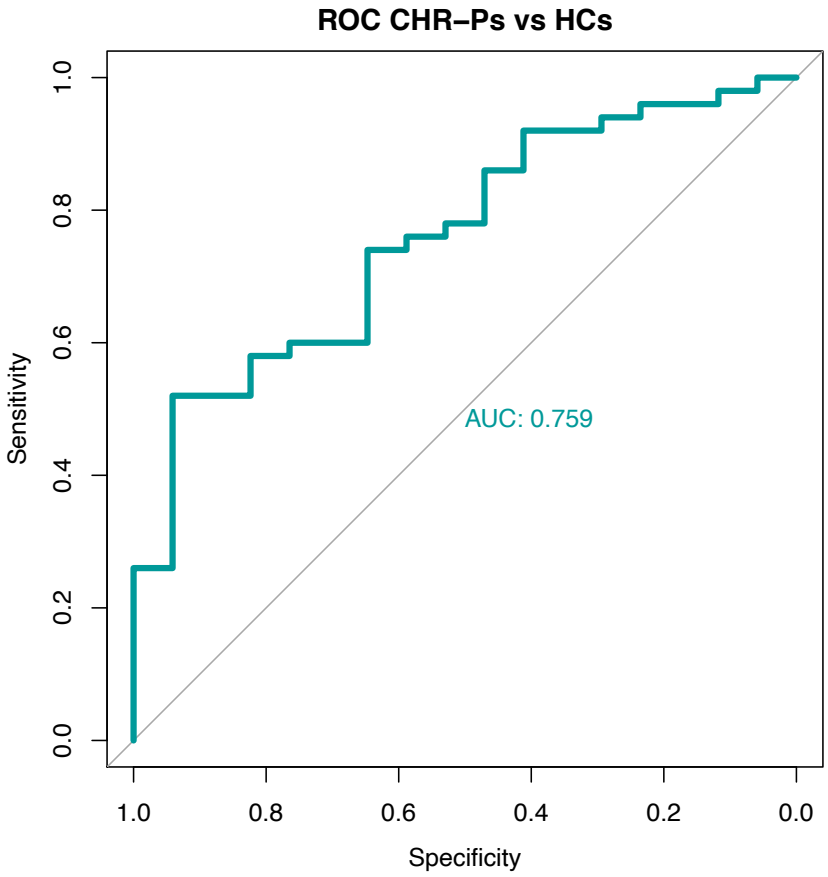

Figure 7 ROC curve for the logistic regression model for temporal uncorrected data discriminating CHR-Ps vs HCs.

Bootstrapping results showed poor model accuracy (AUC 95% CI = 0.45-0.67). I/P speech remained a significant predictor (95% CI = 0.004-0.052), while years of education (95% CI = -0.548, 0.031) and gender (95% CI =-1.403, 2.08) resulted non-significant predictors.

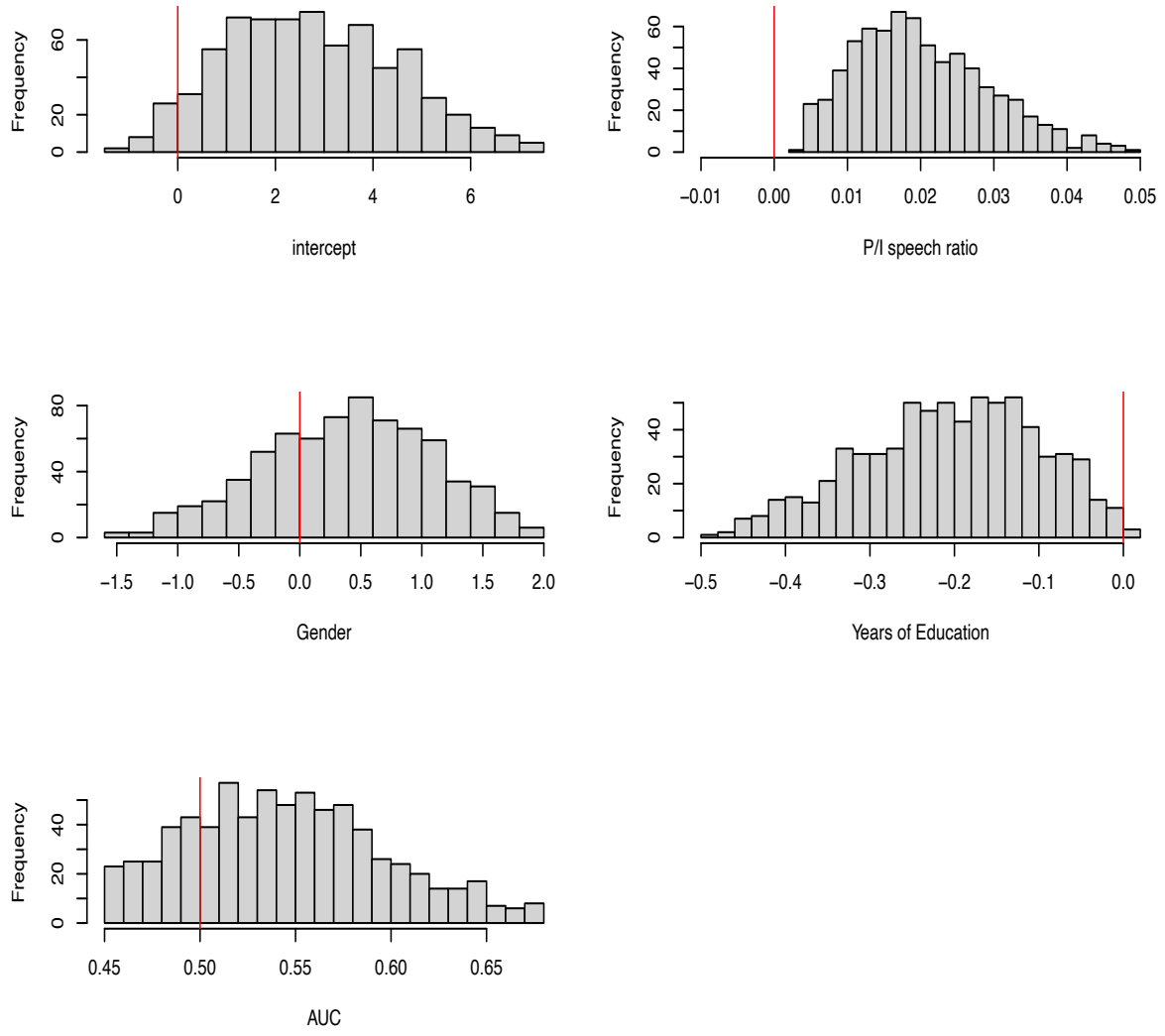

Figure 8 Histograms of the bootstrapping results for the binary logistic regression model of temporal uncorrected acoustic variables determining CHR-Ps and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

Binary Logistic Regression Model for prediction of Group Status from temporal corrected data are depicted in *Table 10*.

Table 10 Binary Logistic Regression Models from temporal corrected variables for prediction of Group Status

Model for prediction of CHR-P vs. CHR-N status					
Variable	***B***	**SE**	**Wald**	**p-value**	**OR (95% CI)**

intercept	0.89	0.29		.002**	2.43(1.41- 4.404)
Average Syllable Duration	-13.05	6.54	4.33	.038*	0.000001(5.3e-13- 0.10)
Average Pause Duration	-2.78	1.21	5.26	.022*	0.062(0.004-0.54)
Total Interview Time	-0.042	0.029	2.20	.138	0.959(0.901-1.009)
Model for prediction of CHR-P vs. HC status					
intercept	3.96	1.47		**.**007**	52.56(3.25 - 1168)
Years of Education	-0.12	0.09	4.11	.043*	0.83(0.69 - 0.99)
Model for prediction of CHR-N vs. HC status					
intercept	0.91	0.53		0.086*	2.47(0.98 - 8.18)
Average Syllable Duration	37.14	17.88	4.32	0.038*	1.34e+16(670.49 - 5.26e+33)
I/P Speech	-0.027	0.011	6.08	0.014*	0.972(0.948- 1.0)

Note: CHR-P, clinical high-risk; CHR-N, clinical low risk; HC, healthy controls; *B,* beta coefficient; SE, standard error; OR, odds ratio; 95% CI, 95% confidence interval. p < .05*; p < .01**; p< .001***

The model obtained for CHR-Ps versus CHR-N accounted for 28% of the variance (NR^2 = 0.283) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.6). This model included two significant predictors, average syllable duration (β = -13.05, OR = 0.000001, 95% CI = 5.3e-13-0.10, p = .038) and average pause duration (β = -2.78, OR = 0.062, 95% CI = 0.004-0.54, p = .022) as well as a nonsignificant predictor, total interview time (β = -0.42, OR = 0.959, 95% CI = 0.901-1.009, p = .138). No source of multicollinearity was detected among the final predictors (VIF: 1.004-1.007). The ROC curve for the temporal corrected model predicting group status for CHR-Ps versus CHR-Ns is shown in *Figure 8.* The discriminative ability of the model was acceptable, with an overall accuracy of 0.77 (AUC).

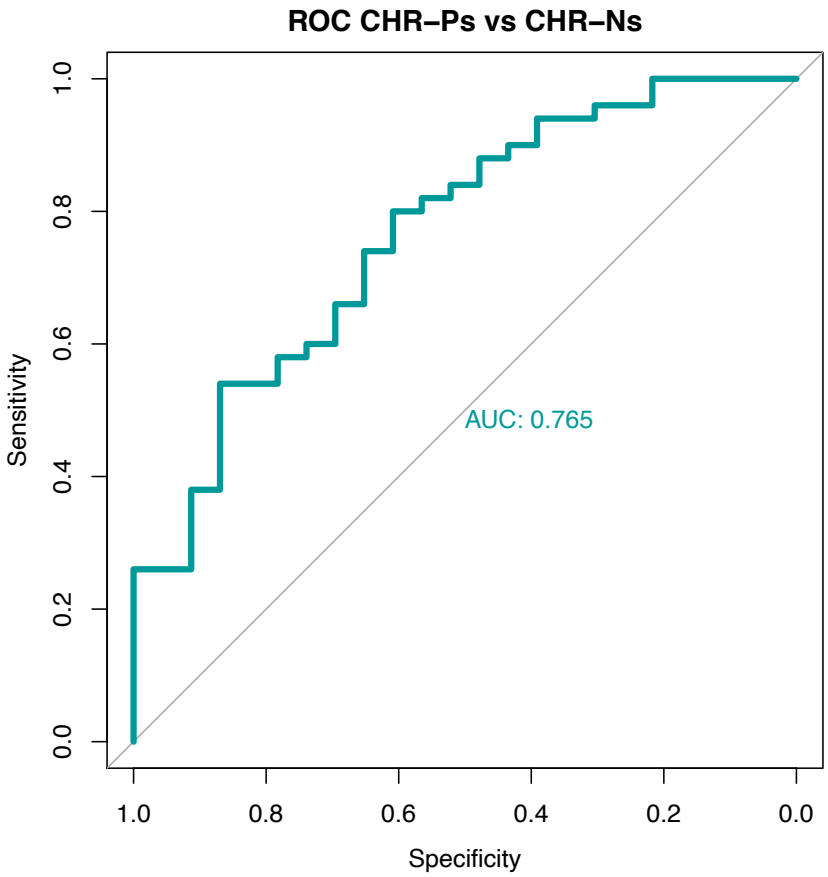

Figure 9 ROC curve for the logistic regression model for temporal corrected data discriminating CHR-Ps vs CHR-Ns

Bootstrapping (*Figure 9*) revealed poor model accuracy, given that bootstrapped AUC included values below chance-performance (AUC, *95% CI* = 0.45, 0.66). However, average syllable duration (*95% CI* = -31.7, -2.13) and average pause duration (*95% CI* = -5.85, -0.52) remained significant, total interview time (*95% CI* = -0.11-0.01) was not significant after bootstrapping.

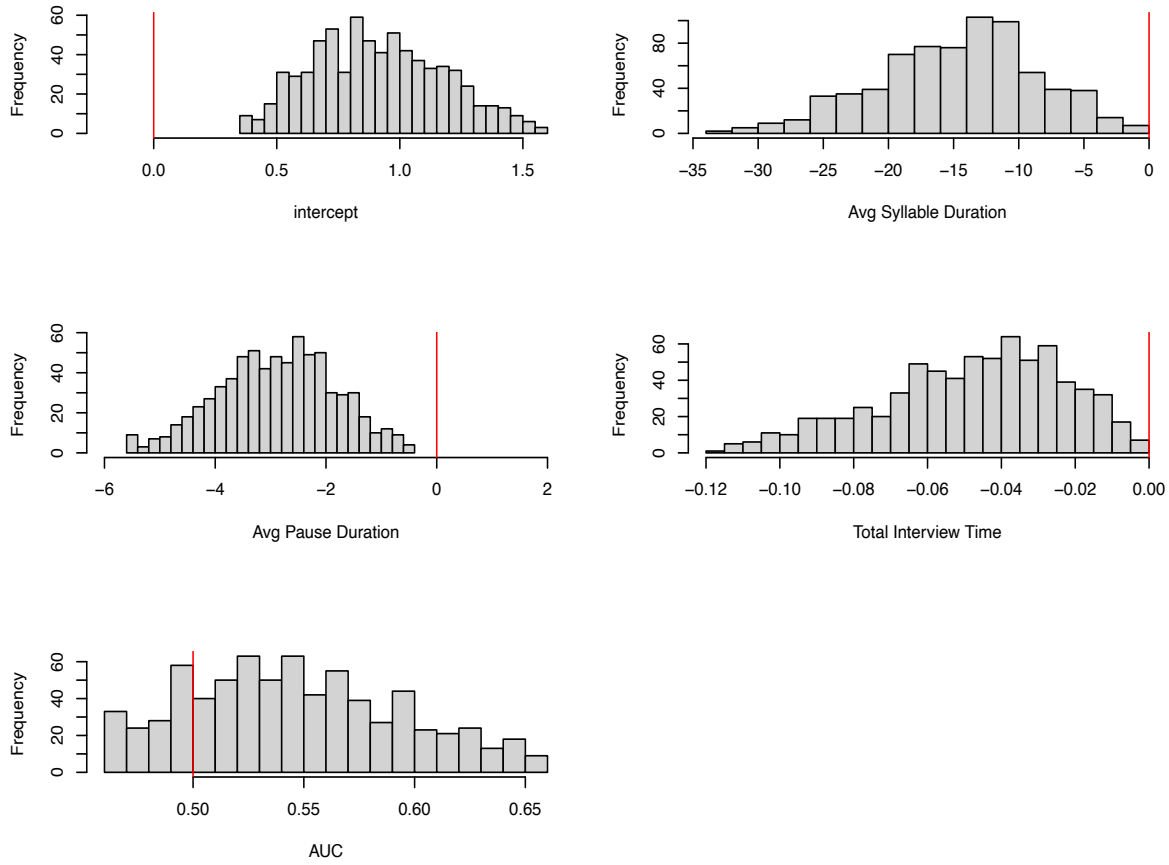

Figure 10 Histograms of the bootstrapping results for the binary logistic regression model of temporal corrected acoustic variables determining CHR-Ps and CHR-Ns. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The temporal model obtained for identification of groups status between CHR-N and HC individuals accounted for 47% of the variance (NR^2 = 0.468) based on average syllable duration (β = 37.14, OR = 1.3e+16, 95% CI = 670.49-5.26e+33, p = .038) and I/P speech ratio (β = -0.027, OR = 0.972, 95% CI = 0.948-1, p = .014). The Hosmer–Lemeshow test for goodness-of-fit was not significant (p = 0.55). Moreover, correlations between predictors indicated lack of multicollinearity (VIF: 1.206). The ROC curve for the temporal corrected model predicting group status for CHR-Ns versus HCs is shown in *Figure 10.* The AUC revealed acceptable discriminative ability (AUC = 0.8).

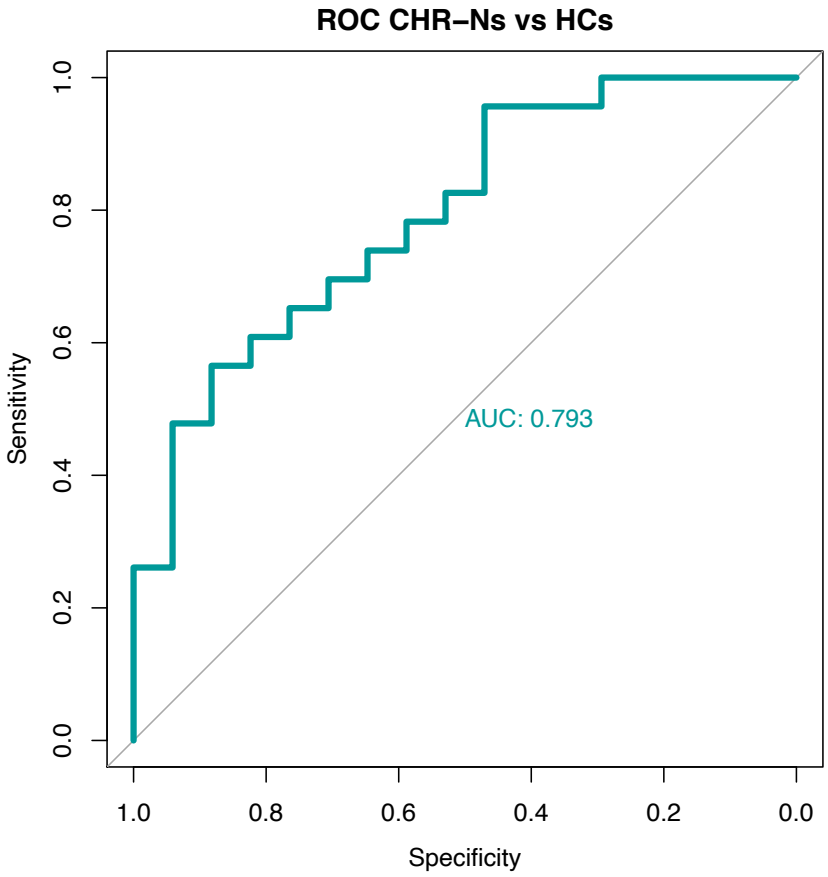

Figure 11 ROC curve for the logistic regression model for temporal corrected data discriminating CHR-Ns vs HCs.

Following bootstrapping, model accuracy included values below chance performance (AUC, 95% CI = 0.45, 0.70), however average syllable duration (*95% CI* = 9.11, 103.1) and I/P speech (*95% CI* = -0.07, -0.01) remained significant (*Figure 11*).

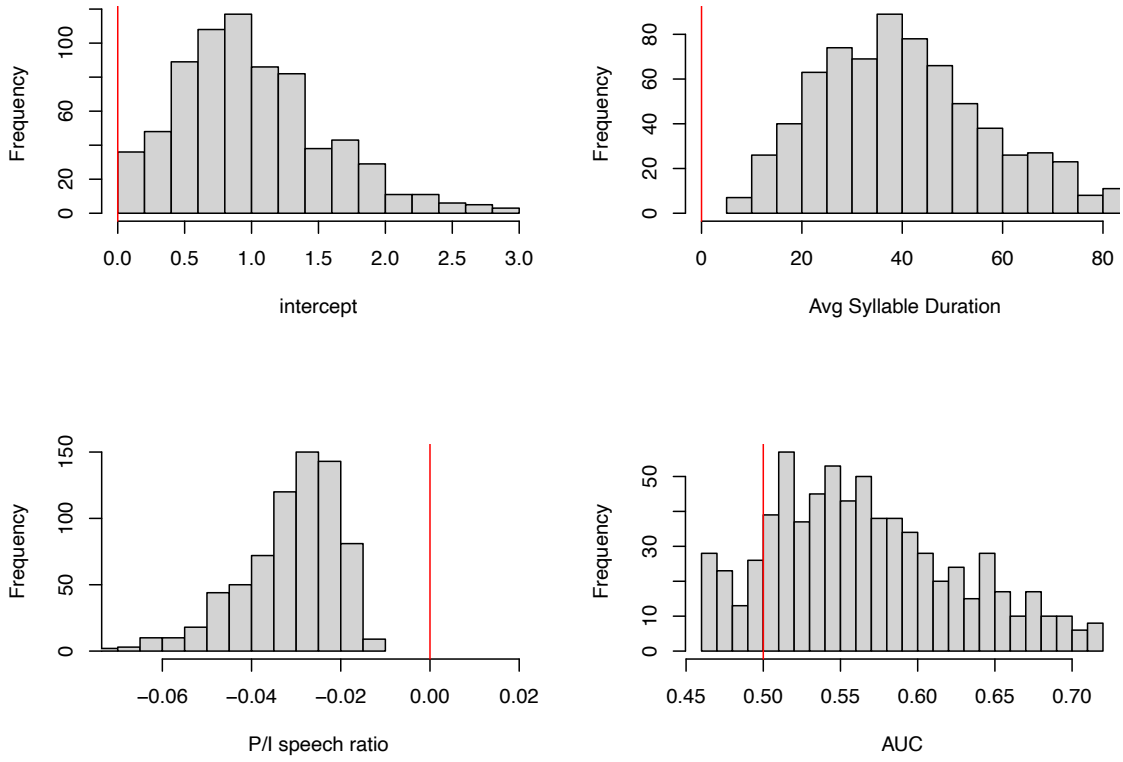

Figure 12 Histograms of the bootstrapping results for the binary logistic regression model of temporal corrected acoustic variables determining CHR-Ns and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The model for detection of group status from temporal corrected data between CHR-P and HC revealed poor fit and accounted for only 9% of the variance (NR^2 = 0.091) based solely on years of education (β = -0.12, OR = 0.83, 95% CI = 0.69 - 0.99, p = .043), while none of the temporal variables were present in the final model. The Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.34). The ROC curve for the temporal corrected model predicting group status for CHR-Ps versus HCs is shown in *Figure 12*. The AUC revealed poor discriminative ability (AUC = 0.63).

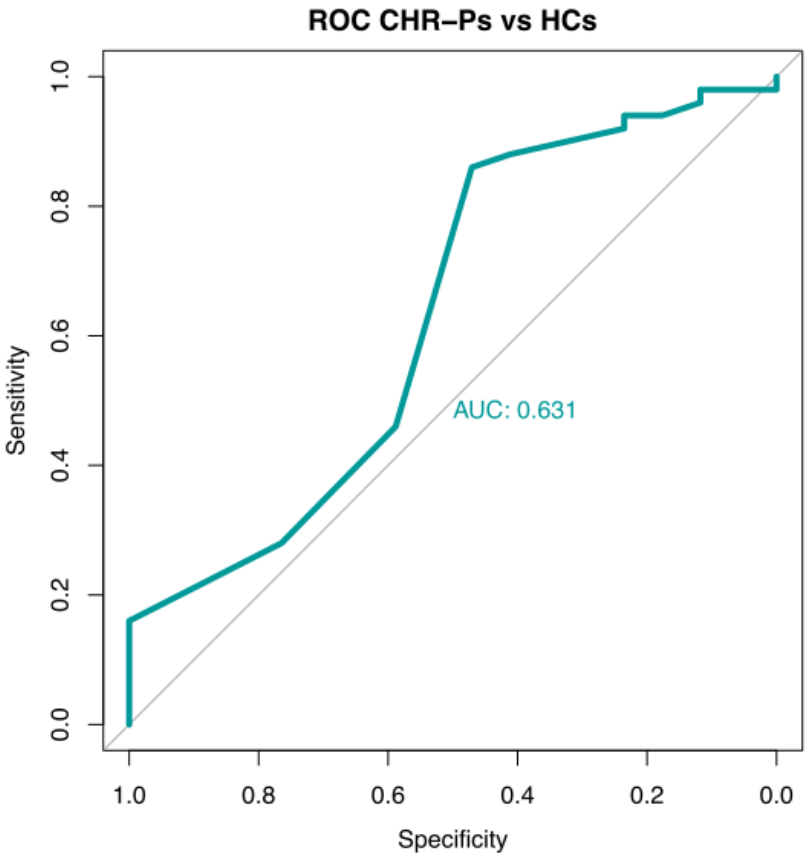

Figure 13 ROC curve for the logistic regression model for temporal corrected data discriminating CHR-Ps vs HCs.

Bootstrapping results revealed once again model accuracy below chance performance (AUC, 95% CI = 0.42, 0.67), but the only predictor of years of education remained significant (*95% CI* = -0.433, -0.008) (*Figure 13*).

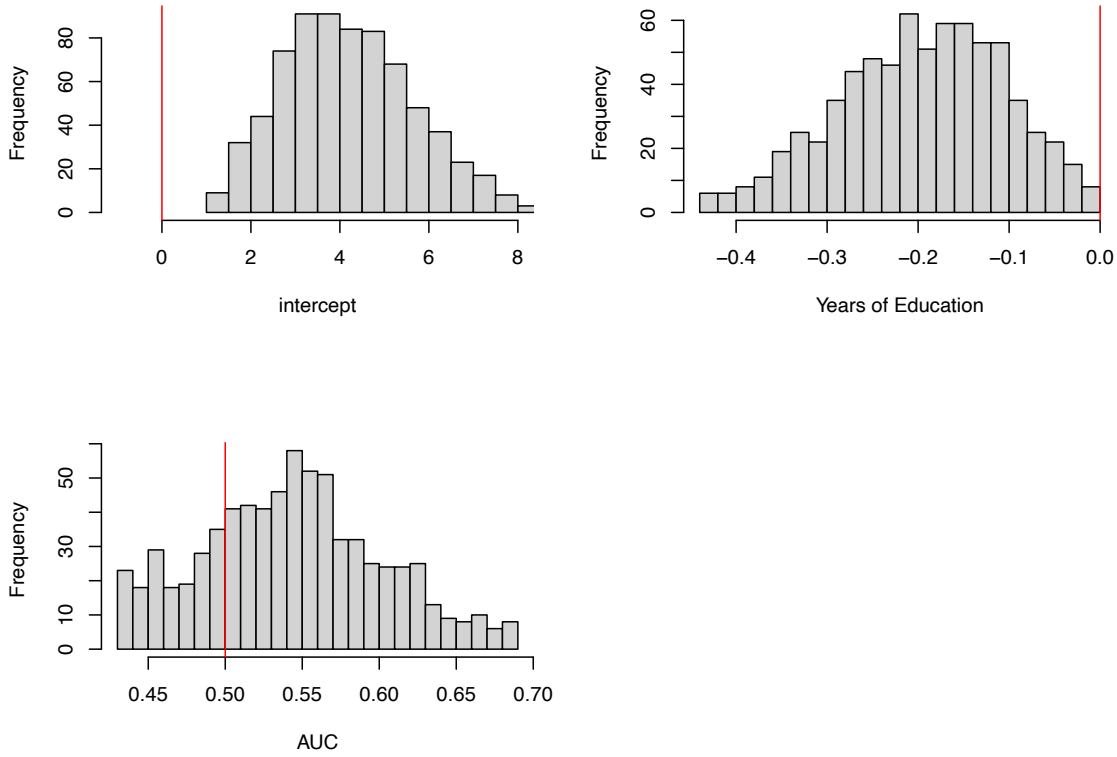

Figure 14 Histograms of the bootstrapping results for the binary logistic regression model of temporal corrected acoustic variables determining CHR-Ps and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

Table 11 shows the regression model results for the identification of group status from the prosodic uncorrected dataset.

Table 11 *Binary Logistic Regression Models from prosodic uncorrected variables for prediction of Group Status*

Model for prediction of CHR-P vs. CHR-N status					
Variable	***B***	**SE**	**Wald**	**p-value**	**OR (95% CI)**
intercept	3.30	1.25		.008**	27.11(2.59 - 378.23)
Years of Education	-0.16	0.08	4.32	.038*	0.85(0.72-0.99)
Model for prediction of CHR-Ps vs. HCs status					

intercept	3.96	1.47		.007**	52.56(3.25 – 1168)
Years of Education	-0.16	0.08	4.11	.042*	0.83(0.689- 0.99)

Note: CHR-P, clinical high-risk; CHR-N, clinical low risk; HC, Healthy controls; *B,* beta coefficient; SE, standard error; OR, odds ratio; 95% CI, 95% confidence interval. p < .05*; p < .01**; p< .001***

The model obtained for CHR-Ps versus CHR-N presented poor fit as it accounted for 0.8% of the variance (NR^2 = 0.008), while the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.88). The model following backwards selection included Voice break percentage, percentage of unvoiced breaks as well as years of education, however the two prosodic variables indicated the presence of multicollinearity (VIF: 24.87-24.87), therefore the final model included solely Years of education (VIF: 1.001). Years of education consisted of a significant predictor (β = -0.16, OR = 0.85, 95% CI = 0.72-0.99, p = .038), indicating that individuals with more years of education had a lower likelihood of falling within the CHR-P group than CHR-Ns. The ROC curve for the prosodic uncorrected model predicting group status for CHR-Ps versus CHR-Ns is shown in *Figure 13.* The AUC revealed poor discriminative ability (AUC = 0.61).

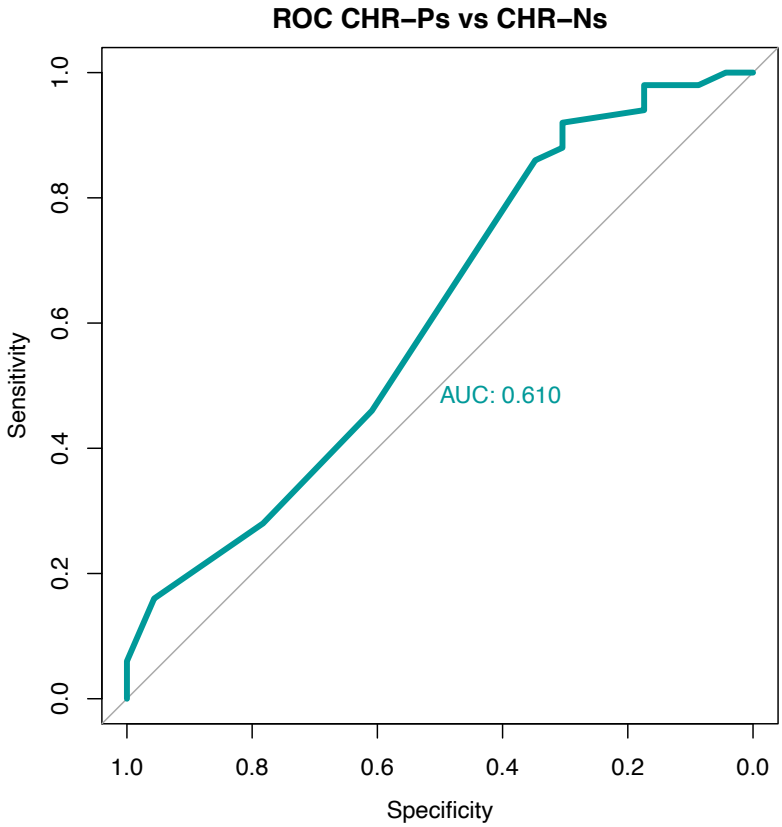

Figure 15 ROC curve for the logistic regression model for prosodic uncorrected data discriminating CHR-Ps vs CHR-Ns

Bootstrapping analysis reported values below chance level (AUC, *95% CI* = 0.41, 0.65) and years of education (*95% CI* = -0.34-0.0002) was not significant after bootstrapping (*Figure 15*).

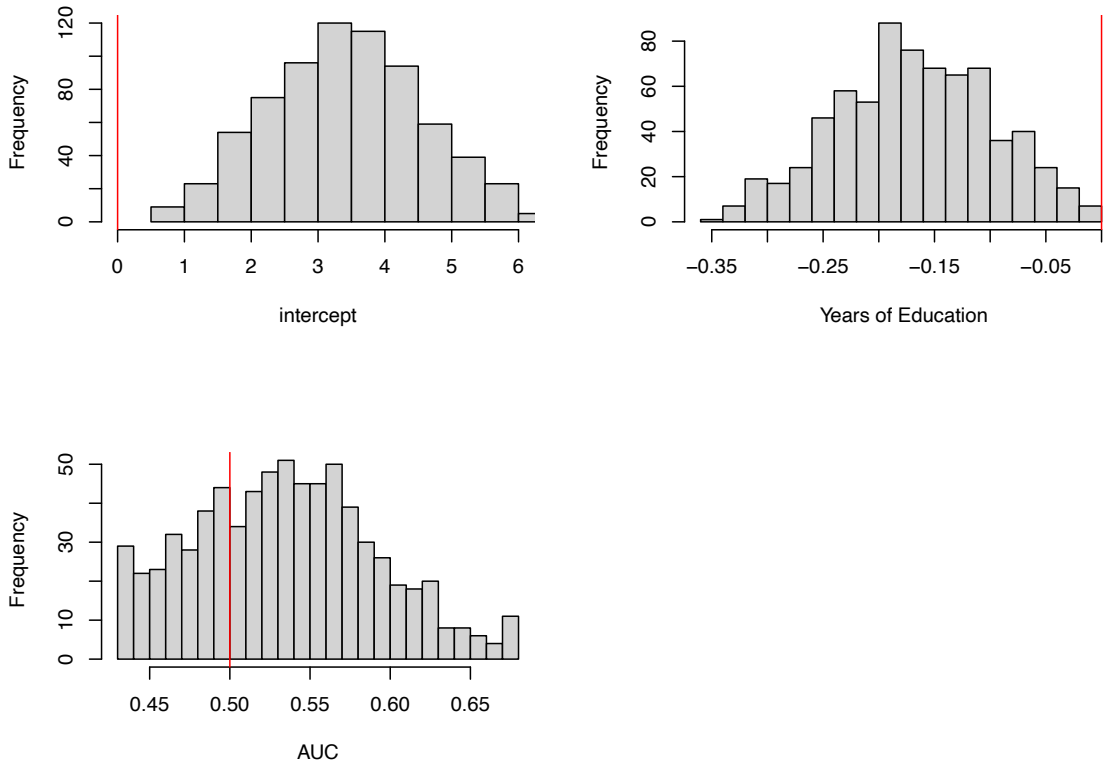

Figure 16 Histograms of the bootstrapping results for the binary logistic regression model of prosodic uncorrected acoustic variables determining CHR-Ps and CHR-Ns. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The prosodic model obtained for CHR-N and HC group status presented two predictors with high multicollinearity values (VIF: 14.58), which were excluded and thus resulted in a final null model, indicating that none of the prosodic variables could discern group membership between CHR-N and HC participants.

For uncorrected prosodic variables, the model obtained to determine group status between CHR-Ps and HCs accounted for 0.9% of the variance (NR^2= 0.009) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.34) and similarly to the model for CHR-Ps vs CHR-Ns identification, it included solely years of education as a predictor (β = -0.16, OR = 0.83, 95% CI = 0.689- 0.99, p = .042). The ROC curve for the prosodic uncorrected model predicting group status for CHR-Ps versus HCs is shown in *Figure 16.* The AUC revealed poor discriminative ability (AUC = 0.63).

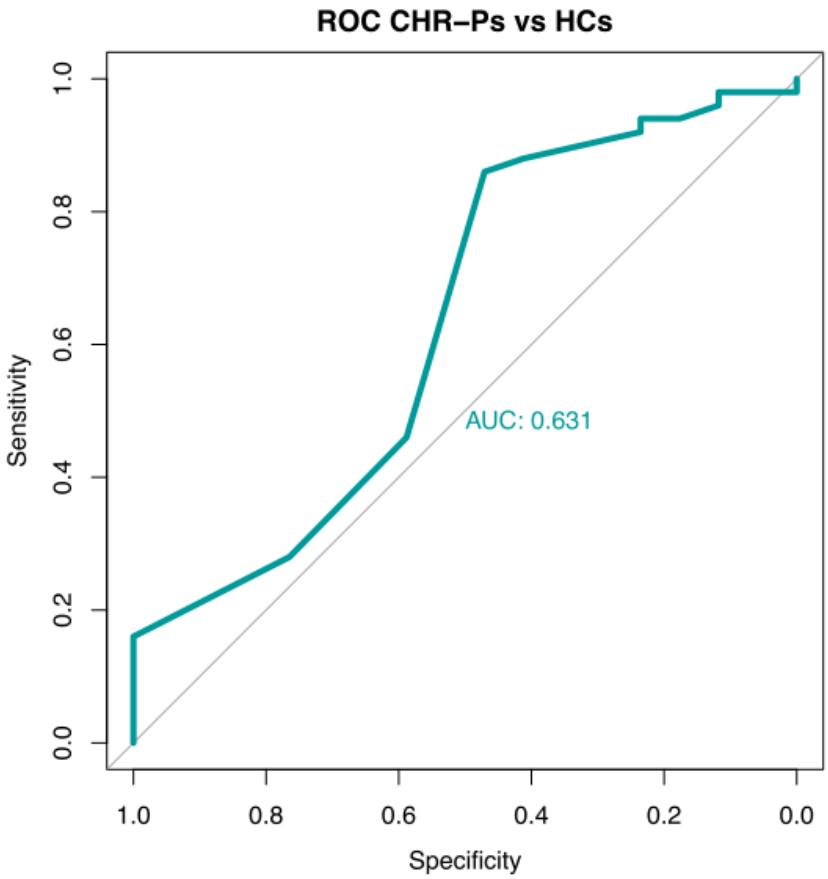

Figure 17 ROC curve for the logistic regression model for prosodic uncorrected data discriminating CHR-Ps vs HCs.

Years of education remailed a significant predictor following bootstrapping (*95% CI* = -0.44,-0.0008) while the model obtained modest diagnostic accuracy (AUC, 95% CI = 0.43, 0.68) (*Figure 17*).

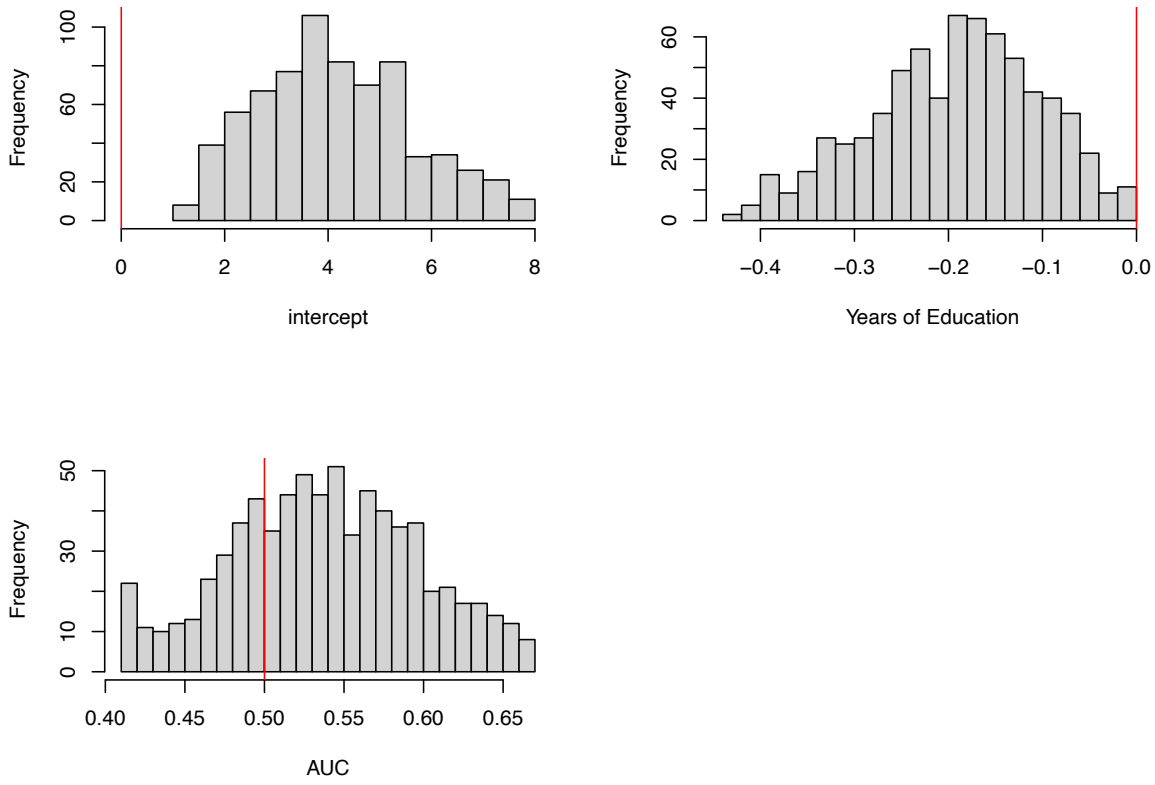

Figure 18 Histograms of the bootstrapping results for the binary logistic regression model of prosodic uncorrected acoustic variables determining CHR-Ps and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

Binary Logistic Regression Model for prediction of Group Status from prosodic corrected data are depicted in *Table 12.*

Table 12 Binary Logistic Regression Models from prosodic corrected variables for prediction of Group Status

Model for prediction of CHR-P vs. CHR-N status					
Variable	***B***	**SE**	**Wald**	**p-value**	**OR (95% CI)**
intercept	4.20	1.44		.004**	66.65(4.59-1443)
Years of Education	-0.20	0.08	5.35	.021*	0.820.6818124 - 0.96)

Unvoiced frames (%)	-0.08	0.03	6.29	.012*	0.93(0.88 - 0.98)
25th pct pitch	-0.02	0.14	2.75	.097	0.80(0.53 - 0.98)
Model for prediction of CHR-P vs. HC status					
intercept	4.82	1.62		.003**	124.09(5.98 – 3739)
Years of Education	-0.23	0.10	5.60	.018*	0.8 (0.65 - 0.96)
Unvoiced Frames	0.06	0.04	2.18	.140	1.06(0.98 - 1.16)
Model for prediction of CHR-N vs. HC status					
intercept	0.22	1.69		.895	1.25(0.042 -35.69)
Age	0.003	0.08	0.002	.962	1.0(.86-1.167)

Note: CHR-P, clinical high-risk; CHR-N, clinical low risk; HC, healthy controls; *B,* beta coefficient; SE, standard error; OR, odds ratio; 95% CI, 95% confidence interval. p < .05*; p < .01**; p< .001***

For the prosodic dataset corrected by interview duration, the model obtained for CHR-Ps versus CHR-Ns accounted for 27% of the variance (NR^2 = 0.265) and the Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.66).This model included two prosodic predictors, unvoiced frames (β = -0.08, OR = 0.98, 95% CI = 0.88, 0.98, p = .012) and the 25th pitch percentile (β = -0.02, OR = 0.80, 95% CI = 0.53-0.98, p = .097) with negative beta coefficients, indicating higher pitch, unvoiced frames percentage reflecting lower likelihood of falling under the CHR-P group. An additional predictor consisted of years of education (β = -0.2, OR = 0.82, 95% CI = 0.68-0.96, p = .021). Correlation values among predictors revealed no source of multicollinearity within the model (VIF: 1.03-1.05). The ROC curve for the prosodic corrected model predicting group status for CHR-Ps versus CHR-Ns is shown in *Figure 18.* The AUC revealed adequate discriminative ability (AUC = 0.77).

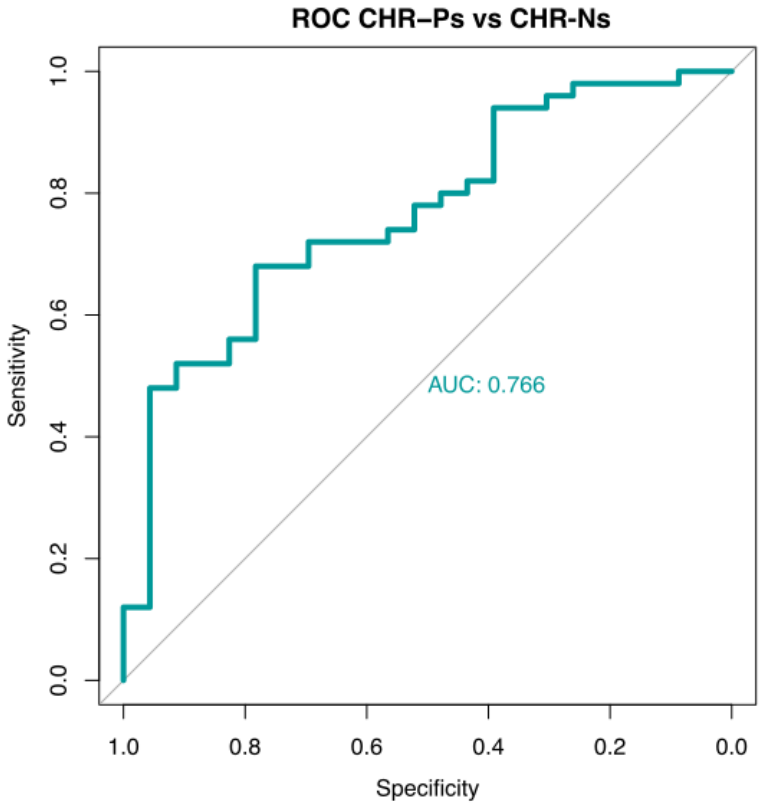

Figure 19 ROC curve for the logistic regression model for prosodic corrected data discriminating CHR-Ps vs CHR-Ns.

Bootstrapping results (*Figure 19*) showed poor model performance (AUC, *95% CI* = 0.46, 0.66) and while unvoiced frames (*95% CI* = -0.15, -0.028) remained significant, the twenty-fifth percentile of pitch variability was not significant (*95% CI* = -0.83, 0.042).

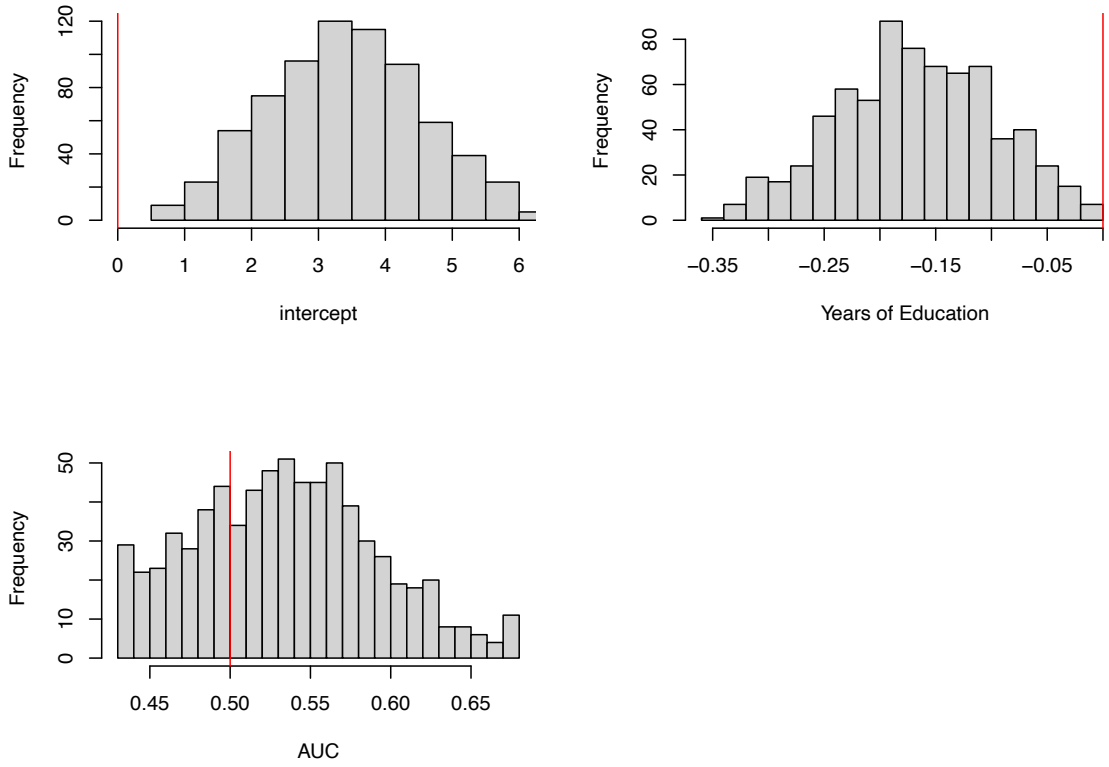

Figure 20 Histograms of the bootstrapping results for the binary logistic regression model of prosodic corrected acoustic variables determining CHR-Ps and CHR-Ns. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The model for prediction of group status between CHR-Ns versus HCs revealed extremely poor fit, accounting for 0.0007% of the variance (NR^2 = 0.00007) based on a single non-significant predictor of age (β = 0.003, OR = 1, 95% CI =0.86,1.17, p = .96). The ROC curve for the prosodic corrected model predicting group status for CHR-Ns versus HCs is shown in *Figure 20.* The AUC revealed poor discriminative ability (AUC = 0.59).

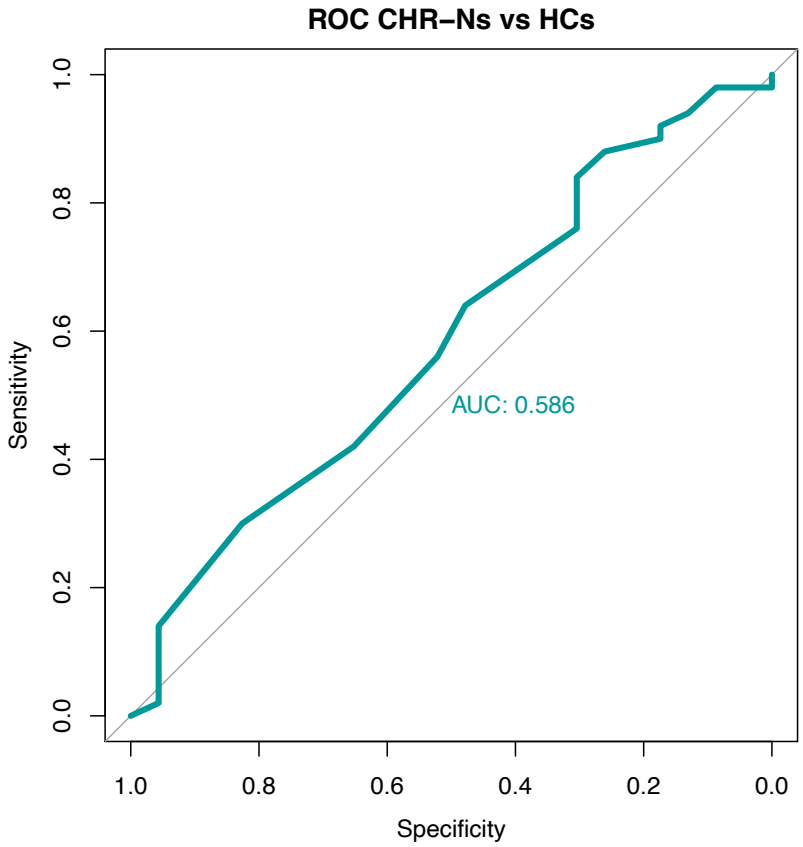

Figure 21 ROC curve for the logistic regression model for prosodic corrected data discriminating CHR-Ns vs HCs.

Following bootstrapping (*Figure 21*), model accuracy performed below chance performance (AUC, 95% CI = 0.43, 0.701 and years of education remained not significant (*95% CI* = -0.19, 0.17).

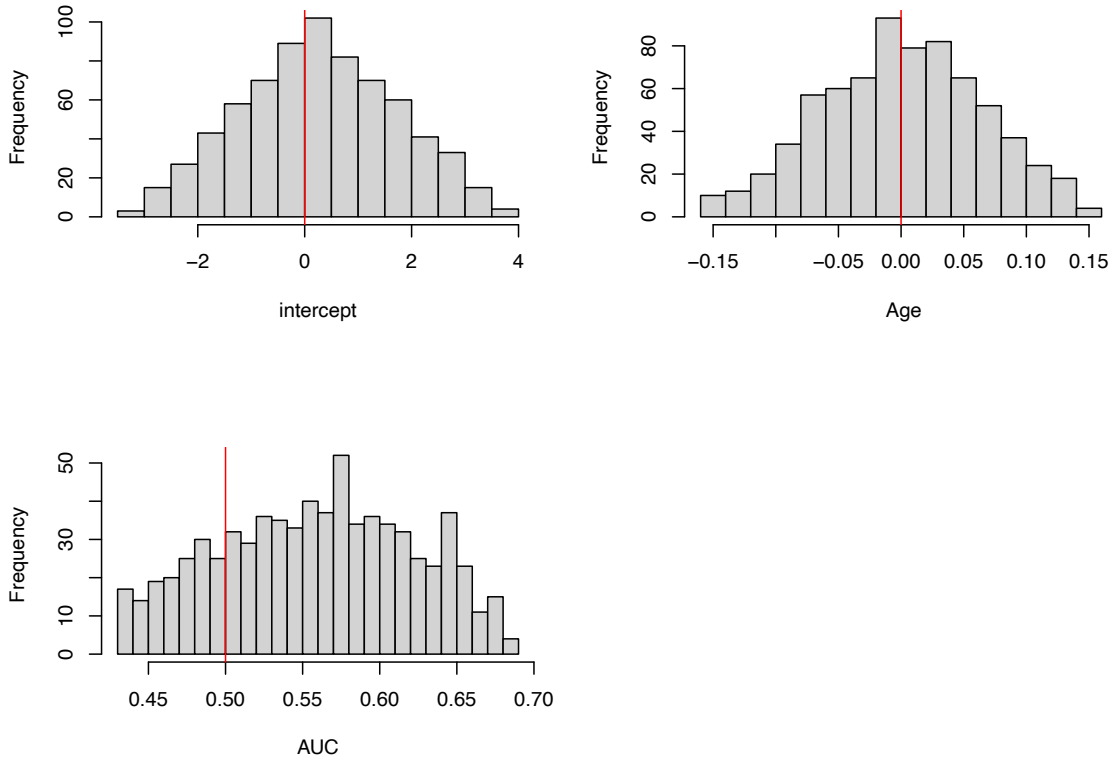

Figure 22 Histograms of the bootstrapping results for the binary logistic regression model of prosodic corrected acoustic variables determining CHR-Ns and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

The model for detection of group status from prosodic data corrected by interview duration between CHR-P and HCs, accounted for only 14% of the variance (NR^2 = 0.14) based on years of education (β = -0.23, OR = 0.8, 95% CI = 0.65 - 0.96, p = 0.018) and the non-significant predictor, percentage of unvoiced frames (β = 0.06, OR = 1.06, 95% CI = 0.98, 1.16, p = 0.140). No source of multicollinearity was detected between the two predictors (VIF: 1.14). The Hosmer–Lemeshow test for goodness-of-fit was non-significant (p = 0.56). The ROC curve for the prosodic corrected model predicting group status for CHR-Ps versus HCs is shown in *Figure 22.* The AUC revealed poor discriminative ability (AUC = 0.63).

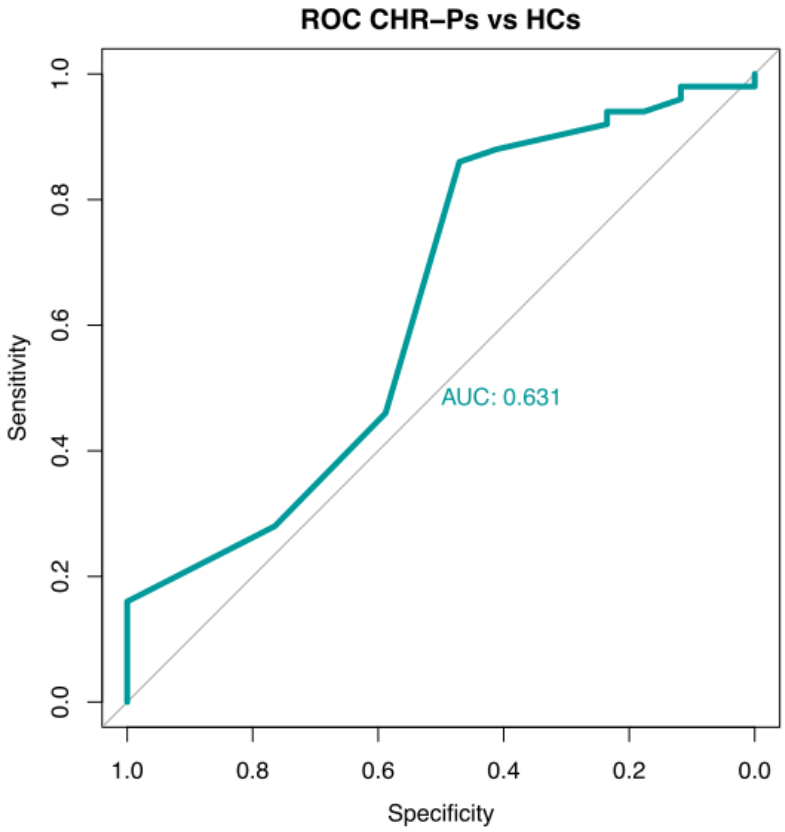

Figure 23 ROC curve for the logistic regression model for prosodic corrected data discriminating CHR-Ns vs HCs

Bootstrapping results (*Figure 23*) revealed once again model accuracy below chance performance (AUC, 95% CI = 0.46, 0.68), with years of education maintaining significance (*95% CI* = -0.48, -0.022) and percentage of unvoiced frames remaining not significant *95% CI* = -0.01, -0.16).

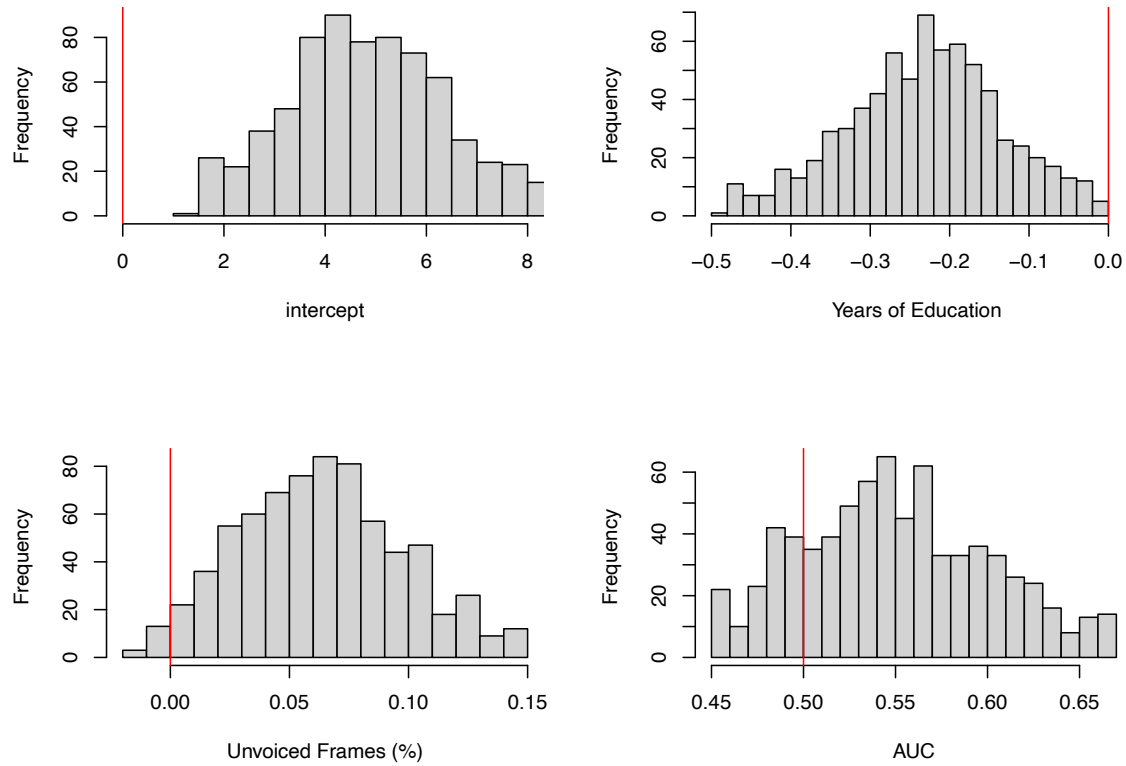

Figure 24 Histograms of the bootstrapping results for the binary logistic regression model of prosodic corrected acoustic variables determining CHR-Ps and HCs. Red vertical lines at value 0 on the x-axis indicate whether the predictors were significant (values below or above 0) or were not significant (values including 0). The red vertical line above 0.5 for AUC values shows chance level.

3.5. Correlations.

We computed Kendall's Tau Coefficient Correlations between each acoustic variable corrected by interview length with baseline clinical and functional measures within the CHR-P group. Significant correlations were obtained between the CAARMS total severity at baseline and the corrected prosodic variables of 25th pitch percentile (τb = -0.27, p = .007) and 5th pitch percentile (τb = -0.29, p = .003), indicating that individuals with higher overall CAARMS severity at baseline presented lower 5th and 25th pitch percentile scores. Moreover, the total SPI-A severity at baseline was negatively correlated with the same prosodic indices of 25th pitch percentile (τb = -0.26, p = .009) and 5th pitch percentile (τb = -0.25, p = .011), suggesting that individuals with higher SPI-A symptom severity had lower pitch values. Significant correlations were observed for most of the CAARMS subscales and prosodic features. The Unusual Thought Content (UTC) item of the CAARMS was significantly associated with 5th pitch percentile (τb = -0.23, p = .032) and 25th pitch

percentile (τb = -0.22, p = .037), once again indicating that individuals with higher scores had lower pitch values. The Perceptual Abnormality (PA) item of the CAARMS presented a negative association with the mean value of glottal pulse period (τb = -0.22, p = .043) and similarly the Disorganised speech (DS) subscale was negatively linked with the mean value of glottal pulse period (τb = -0.22, p = .047). The DS item of the CAARMS positive scale showed negative correlations with a number of pitch variability determinants including pitch skewness (τb = -0.23, p = .033), kurtosis (τb = -0.25, p = .024), 5th (τb = -0.25, p = .022) and 95th (τb = -0.27, p = .012) percentiles of pitch, indicating that individuals with higher disorganised speech values presented less skewed values with less extreme values in the distribution and overall lower pitch percentiles. However, none of the observed significant correlations reached significance following FDR multiple comparison corrections.

No significant correlations were found between any of the acoustic variables and functioning scores measured with the GAF, GF: Social and GF: Role scales. Furthermore, no significant correlation was obtained for any of the temporal features and the clinical/functional measures examined. All non-significant correlations are depicted in *Appendix Table B.1* and *Appendix Table B.2*.

Table 13 Correlations between language variables and baseline clinical/functional measures

Clinical/functional measures	***Acoustic features***	**Kendall's τ**	***Z-scores***	***p-values***	***FDR correction***
CAARMS total severity	**Pitch 25th pct**	-0.27	-2.69	.007**	.097
	Pitch 5th pct	-0.29	-2.94	.003*	.993
SPI-A	**Pitch 25th pct**	-0.26	-2.60	.009**	.168
	Pitch 5th pct	-0.25	-2.54	.011*	.168
CAARMS Subscales:					
NBI	**Pitch 5th pct**	-0.23	-2.15	.032*	.561
	Pitch 25th pct	-0.22	-2.08	.037*	.561

PA	**Glottal Pulses (Mean)**	-0.22	-2.02	0.043*	.870
DS	**Glottal Pulses (Mean)**	-0.216	-1.99	.047*	.281
	Pitch Skewness	-0.23	-2.13	.033*	.248
	Pitch Kurtosis	-0.25	-2.26	.024*	.240
	Pitch 5th	-0.25	-2.29	.022*	.240
	Pitch 95th	0.27	2.51	.012*	.240

Abbreviations: CAARMS, Comprehensive Assessment of At-Risk Mental States; UTC, Unusual Though Content; PA, Perceptual Abnormalities; DS, Disorganised Speech; GF: Social, Global Functioning: Social; pct, percentile; ST, semitones; SD, NHR, noise to harmonics ratio; HNR, harmonics to noise ratio. p < .05*; p < .01**; p < .001***

3.6. Effects of medication status.

Diagnosis (CHR-Ps vs CHR-Ns) remained a significant predictor of the acoustic variables after controlling for ADMs-status as medication status did not influence any of the speech parameters (see *Table 14* for speech parameters that indicated a significant effect of diagnosis).

Table 14 *Linear regressions on the influence of ADMs on speech parameters between CHR-Ps and CHR-Ns*

Acoustic variables	Predictors	β	S.E.	t	*p*
Speech rate	intercept	0.20	0.12	1.74	0.087
	Diagnosis (CHR_N)	-0.56	0.20	-2.82	0.007 **
	medication	-0.12	0.19	-0.64	0.524
	Diagnosis*medication	0.22	0.35	0.63	0.532
Mean length of runs	intercept	-0.03	0.015	-1.93	0.058
	Diagnosis (CHR_N)	0.07	0.026	2.69	0.009 **
	medication	0.03	0.024	1.27	0.209

	Diagnosis*medication	-0.05	0.045	-1.07	0.287
Pause time (%; adjusted)	intercept	-3.557	1.52	-2.34	0.022 *
	Diagnosis (CHR_N)	6.88	2.61	2.64	0.010 **
	medication	4.52	2.47	1.83	0.072
	Diagnosis*medication	-2.48	4.57	-0.54	0.589
Unvoiced frames (%)	intercept	-2.18	1.66	-1.31	0.194
	Diagnosis (CHR_N)	8.31	2.85	2.91	0.005 **
	medication	2.60	2.70	0.96	0.340
	Diagnosis*medication	-6.86	4.99	-1.37	0.174

P-values: * $p < .05$; ** $p < .01$; *** $p < .001$

Abbreviations: CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; *AMDs*, antidepressant medication; adjusted, relative to the total interview duration; %, percentage; β, beta coefficient; S.E, standard error; t, t statistics; p, p-value.

4 Discussion.

4.1 Summary of the results.

We investigated temporal and prosodic features of speech in CHR-Ps to identify whether acoustic speech parameters were impaired in the CHR-P cohort compared to controls and could constitute a biomarker for early detection and diagnosis.

Surprisingly, the group comparison effects emerging for the temporal features revealed that CHR-Ps were characterised by faster speech rate, more efficient speech production (measured by mean length of runs) and less time spent pausing compared to HCs. Moreover, the ratio of participant and interviewer speech was larger in the HC group compared to both CHR-P and CHR-Ns participants. The present analysis revealed that the group differences for temporal and prosodic features were largely robust to the interview duration effects. Apart from the P/I speech ratio, the temporal findings were similar for data before and after accounting for interview duration effects. The prosodic analysis revealed only one significant index across groups, which was unvoiced frames, lower in CHR-Ps compared to CHR-Ns, both for corrected and uncorrected data.

Overall, binary logistic regression models that included temporal features (temporal model) presented higher variance explained and a better discriminant ability between HC and CHR-P prior to validation compared to models including prosodic features (prosodic model). Two temporal models could explain a considerable amount of variance from datasets obtained before and after correcting for interview duration in determining group status of CHR-N and HC (35% and 47% of the variance explained) and CHR-P and CHR-N (28% of the variance explained for both datasets). Most prosodic models could explain a very small portion of the data and the majority of these only included the covariate of years of education or age as a predictor, indicating that the prosodic features did not perform efficiently when distinguishing group status. The only regression model including prosodic features that could explain a considerable portion of the variance consisted of the model comparing CHR-Ps vs CHR-Ns based on corrected prosodic features yielding 27% of the variance explained. However, this result did not survive bootstrapping.

The correlations between temporal and prosodic features with clinical and functional symptom severity revealed significant associations between prosodic features and CAARMS as well as SPI-A overall symptom severity within the CHR-P group. The correlation results indicated that higher scores of positive and basic symptom severities were associated with lower prosodic impairments of pitch variability and instability. Nonetheless,

none of the observed correlations remained significant following multiple comparison corrections.

4.2 Effect of interview duration.

The group comparison results revealed that the acoustic features examined remained largely invariant to the effects of interview duration, characterised by spared acoustic features in CHR-Ps and HCs, but indicating acoustic aberrations in the CHR-N cohort. From visual inspection of scatterplots depicting acoustic metrics before and after correction, it emerged that the relationship between groups remained virtually indistinguishable. Except for P/I speech ratio which presented a significant directional change following interview duration correction. Yet, as mentioned in the results section, the correction of the participant's speech duration but not of the interviewer's speech duration is responsible for the apparent divergence of the P/I speech ratio index before and after correcting for speech duration effects. The group comparison results additionally confirmed that the group differed only marginally after accounting for the speech duration effects, whereby the corrected data indicated for the most part nearly identical effects for the acoustic properties with the addition of the temporal index of P/I speech ratio being significantly higher in CHR-Ns compared to the other two cohorts and the prosodic index of unvoiced frame percentage higher in CHR-Ns compared to CHR-P and HC participants.

These findings are in line with previous results showing that analysing a shorter segment of speech sample or the entire duration of the recording can equally detect mood disorder characteristics from spontaneous speech (Alghowinem *et al.*, 2013). This result is of crucial importance in relation to previous work, given that this is the first study within the bulk of literature to directly investigate whether and to what extent the length of the interview affects the acoustic metrics. Although interview duration has been controlled for as a covariate in previous work investigating acoustic features in CHR-P individuals (Sichlinger *et al.*, 2019) and in ScZ patients (Covington *et al.*, 2012; Cohen *et al.*, 2016, 2020), the majority of studies eliciting speech from clinical interview methods (Tahir *et al.*, 2019; de Boer, Voppel, *et al.*, 2020) have not investigated or accounted for the plausible confound of speech sample duration. This is relevant for studies that, similarly to the present one, have used clinical assessments to analyse speech features, given that it follows from reason that individuals meeting the clinical criteria might obtain longer interviews compared to those participants that do not present the symptoms. Thus, such strong variation in the speech duration across groups would likely yield differences in the temporal and prosodic indices. Nonetheless, the present finding indicates that automated acoustic analysis extracts vocal information is

robust to the variations in speech duration; thus, speech duration did not consist of a confound of the acoustic effects observed. However, to determine the generalisability of this finding, future studies should further investigate the effect of the speech sample duration in separate subject pools.

4.3 Group differences.

The group difference results showed that CHR-P participants were characterised by more efficient speech production and less impaired prosodic features compared to CHR-Ns, while no significant differences for the acoustic variables were observed between CHR-P and HC individuals.

These findings suggest that temporal and prosodic speech parameters may be intact in the CHR-P cohort examined in the present study. These results stand in sharp contrast with previous studies that reported temporal and prosodic deficits in CHR-P participants (Sichlinger et al., 2019; Agurto et al., 2020). However, these acoustic impairments emerged solely for within-group effects, when the acoustic features in CHR-Ps were examined in relation to symptomatology or transition rate, but no between-group effects were shown for the acoustic parameters in the comparison of CHR-P to HC individuals (Sichlinger *et al.*, 2019). The existing findings indicate that, in the at-risk cohort, acoustic abnormalities may only emerge when accounting for the high level of heterogeneity within the CHR-P population (Fusar-Poli, Cappucciati, *et al.*, 2016). Crucially, it is plausible that aberrations in temporal and prosodic features may be specific to the subgroup of CHR-Ps that will later develop psychosis. But when examining the entire CHR-P cohort in comparison to HC participants, the speech features may be rather subtle. We speculate that this pattern might indicate that in CHR-Ps more subtle clinical symptomatology goes hand in hand with the less evident presence of acoustic impairment. This pattern differs from the clearer presence of acoustic aberrations and clinical symptoms in chronic ScZ (Rapcan *et al.*, 2010; Compton *et al.*, 2018; Tahir *et al.*, 2019; de Boer, van Hoogdalem, *et al.*, 2020; De Boer *et al.*, 2021).

However, in the present thesis, individuals with higher symptom severity did not present more impaired acoustic features than CHR-Ps with lower symptomatology, as shown by the correlation results. Thus, the spared acoustic abilities in CHR-Ps may instead be explained by the heterogeneity of psychopathology and disparate comorbidity facets in CHR-Ps (Millman *et al.*, 2019). Indeed, most studies that have examined acoustic characteristics have focused on flat affect although other clinical aspects could lead to voice differences. Psychosis symptoms such as hostility, mania and aspects of thought disorder (such as

tangentiality) are conceptually related to vocal exaggeration including pressured speech in contrast to the deficits in speech production such as flattened vocal affect (Sobin and Alpert, 1999). While vocal exaggeration would increase acoustic ability, flattened vocal affect would result in the reduction of pitch variation, yielding opposite acoustic impairments. The plausible presence of opposing speech impairments (vocal exaggeration vs deficit) in the CHR-P group examined in the present work may have hidden the presence of acoustic impairments and could originate in different mechanisms involved in the emergence of acoustic deficits. Unfortunately, speech impairments such as vocal exaggeration were not investigated in sufficient depth in this study and therefore this argument remains speculative. Future studies may examine the presence of acoustic deficits as well as semantic and syntactic impairments (tangentiality or impaired semantic coherence) to investigate the mechanisms involved in these opposite speech aberrations.

Contrasting evidence was obtained from studies in ScZ patients, that observed prosodic and temporal impairments in ScZ (Martínez-Sánchez *et al.*, 2017; Compton *et al.*, 2018). Compton *et al.* (2018) divided ScZ patients based on the presence or absence of aprosody. Indeed, it has been found that not all psychosis individuals present speech impairments, whilst a small portion of HCs do (Hitczenko, Mittal and Goldrick, 2021), indicating the non-dichotomous nature of acoustic features between psychosis and healthy individuals. Distinguishing participants with or without speech impairments may have enabled the authors to obtain effects that might have missed in the present study. However, comparing the present results with evidence from ScZ patients is hindered by the heterogeneous nature of ScZ and CHR-P populations.

4.3.1 Acoustic impairments in CHR-Ns.

We observed greater impairment in the acoustic features in CHR-N participants compared to CHR-Ps. This finding is in contrast with prior evidence of overlapping acoustic impairments across different psychopathologies including ScZ (Cummins *et al.*, 2015; Cohen *et al.*, 2020) and also diverges from those studies that have observed psychosis-specific acoustic impairments when compared to autism spectrum disorder and right hemisphere damage (Fusaroli *et al.*, 2017, 2019). Contrarily to prior work, the CHR-N group examined in the current study included a range of non-psychotic psychopathologies. If examined as a cluster of disorders is likely to yield widely disparate acoustic impairments within the same group. Indeed, while depressive symptoms relate to flattened pitch variation (Cummins *et al.*, 2015), anxiety-related symptoms have been associated with vocal instability (Mendoza and Carballo, 1998). According with prior evidence of high comorbidity rates observed in CHR-Ps (Millman *et al.*, 2019), the CHR-P and CHR-N

participants in our sample presented a series of non-psychotic psychopathologies. ON one hand, the CHR-P group presented a similar proportion of mood and axiety disorder characterised by 68% of anxiety disorders and 60% of mood disorder. On the other hand, CHR-Ns reported a prevalence of 57% of anxiety disorders, but only 21.7% of mood disorders. Given the greater proportion of anxiety disorders compared to mood disorders in CHR-Ns, it is plausible that the ratio of anxiety disorder in this cohort may have led to the prominence of anxiety-related acoustic impairments, such as vocal instability or aberrant pauses (Mendoza and Carballo, 1998; Cohen *et al.*, 2016). This imbalance in the psychopathologies present in CHR-Ns could have resulted in greater acoustic impairment in CHR-Ns compared to CHR-Ps despite both cohorts presenting high non-psychotic psychopathology rates.

Higher unvoiced-frame percentages were observed in CHR-Ns compared to CHR-Ps and HCs. The role of unvoiced frames effects in psychosis is unclear and only Agurto *et al.* (2020) investigated unvoiced frames in CHR-Ps but found this feature not highly predictive of psychosis transition. Unvoiced frames consist of non-periodic sounds, observed during the pronunciation of consonants (Zhang and Jiang, 2008) and are a measure of vocal instability. Since voice instability indices are associated with stressor-provoked anxiety (Mendoza and Carballo, 1998), unvoiced frames consisting of a voice instability index may in turn be conceptually related to aberrant arousal levels (Cohen *et al.*, 2016). Anxiety disorders consist of the highest comorbid symptom within CHR-P individuals (68% of CHR-Ps) and therefore the higher percentage of unvoiced frames may indicate higher arousal levels in the CHR-N group. Furthermore, unvoiced frames are commonly examined using sustained vowels, during which the presence of unvoiced frames indicates an impairment (Kiss and Vicsi, 2014). In continuous speech, however, silences occur naturally during plosive articulation (Rosen, 1992). Therefore, depending on the participant's word usage, the proportion of plosives will vary across participants, possibly accounting for higher unvoiced frames in CHR-Ns as a spurious effect.

4.4 Prediction of diagnostic accuracy.

Emerging from the regression analysis, most temporal models could account for a considerable amount of variability, specifically for those models defining group status between CHR-N and HC and between CHR-P and CHR-N individuals. All prosodic regression models performed worse than the temporal ones. The regression analysis findings suggest that the temporal models could explain a wider portion of the variance compared to the models including prosodic features. In accordance with the group comparison results, the temporal metrics could better detect group status. Opposite findings emerge from one

previous study in CHR-Ps which investigated the combined contribution of prosodic and temporal features and observed that only prosodic variables including vowel space and vocal instability (Jitter, shimmer) were highly predictive of psychosis transition (Agurto *et al.*, 2020). Yet, the comparability of these two studies is hindered by the different scopes of our analyses. Whilst we assessed group status, Agurto *et al.* (2020) examined psychosis prediction, although only a limited number of CHR-Ps transitioned to psychosis (n = 5) in their dataset (Agurto *et al.*, 2020), suggesting that the predictability of prosodic metrics in determining transition rate should be interpreted with caution. Nonetheless, investigating CHR-Ps trajectories and clinical subtypes and how these relate to acoustic features may provide more promising acoustic effects given the heterogeneous manifestation observed in CHR-Ps (Fusar-Poli *et al.*, 2015) and therefore holding greater clinical utility than discerning CHR-Ps from HCs. However, the predictability of the temporal and acoustic features examined was not ventured in the present study given the small portion of CHR-Ps from our sample that transitioned to psychosis (n = 4, obtained from 24-month follow up).

Despite some divergence of our regression results with previous work in CHR-Ps, the higher goodness of fit of models including temporal features is in line with previous studies within the ScZ population that examined both temporal and prosodic variables and found only temporal variables as diagnostic of group status (Tahir *et al.*, 2019; de Boer, van Hoogdalem, *et al.*, 2020; de Boer, Voppel, *et al.*, 2020). Nonetheless, there are substantial differences between the ScZ studies and the present one, including the investigation of ScZ patients rather than CHR-Ps, given that these populations are likely to present different acoustic characteristics.

4.4.1 Bootstrapping.

4.4.1.1 Descriptive vs. predictive modelling.

The present bootstrapping results revealed underwhelming accuracy of the models, given that none of the bootstrapped models obtained accuracy values significantly above-chance, despite the discriminant ability of the original models before bootstrapping was acceptable. The divergent accuracy of the regression models before and after bootstrapping, highlights: i) the effect of outliers in the regression models, probably driving the acceptable discriminant ability of the original models before bootstrapping and ii) the difference between descriptive modelling and predictive modelling (Shmueli, 2010). The former holds the purpose of capturing the relationship between the acoustic variables and group status without making any causal or predictive inference, whilst predictive modelling aims to use statistical models

to predict novel or future observations. The crucial difference between these two modelling types is that while descriptive modelling operates at the conceptual level by explaining the construct underlying the model built, predictive modelling operates at the measurable level by generating predictions from the data at hand (Shmueli, 2010). Thus, the apparently promising accuracy obtained by the original models (before bootstrapping) in the present study was then disproved by the poor prediction accuracy of the bootstrapped regressions. Indeed, obtaining acceptable accuracy is rather arduous in predictive modelling because more data is required to obtain lower bias and variance due to the added uncertainty in predicting new data (Shmueli, 2010).

4.4.1.2 Clinical relevance of modelling approaches.

This is particularly important given that much of the evidence that acoustic features can distinguish ScZ or CHR-P individuals from HCs or other psychopathology groups has used machine learning methods without performing any validation procedures (Compton *et al.*, 2018; Sichlinger *et al.*, 2019; Agurto *et al.*, 2020; de Boer, van Hoogdalem, *et al.*, 2020; de Boer, Voppel, *et al.*, 2020) or have conducted validation methods which should be interpret with caution given the relatively small sample size at hand (i.e. k-fold cross-validation on samples below 50 participants; Rapcan *et al.*, 2010; Tahir *et al.*, 2019; Cohen *et al.*, 2020). Therefore, the spurious discrimination ability of the regression models, stress the importance of discerning between modelling types, given that true clinical utility is obtained with predictive modelling.

Importantly, although the original models before validation obtained acceptable goodness of fit and discriminant ability, these models indicated the presence of acoustic impairments in CHR-Ns rather than in CHR-Ps. This emerged from the better fit of models determining group status between CHR-Ns and HCs compared to the poorer fit of the CHR-Ps and HCs model. In addition, the model determining CHR-P vs CHR-N group status was characterised by longer average syllable and pause duration (temporal corrected features) and a greater percentage of unvoiced frames as well as lower 25th pitch percentage (prosodic corrected indices) in CHR-Ns rather than in CHR-Ps. Thus, the presence of acoustic impairments within the CHR-N group, in combination with the poor performance of all models following bootstrapping, seems to indicate the presence of spurious effects in the classification emerging from the CHR-N group, possibly driven by a few subjects leveraging the results (i.e. outliers).

4.4.1.2 The importance of validation methods.

Crucially, an additional aspect differentiating the present study from the previous bulk of literature consists in the validation methods employed. Previous ScZ studies assessing acoustic features utilised machine learning methods on subject pools below 50 (Sichlinger *et al.*, 2019; Tahir *et al.*, 2019; Agurto *et al.*, 2020; Cohen *et al.*, 2020; de Boer, Voppel, *et al.*, 2020). It is commonly assumed that a model requires the number of training samples to be approximately 10 times higher than the degrees of freedom to prevent overfitting (Abu-Mostafa, Magdon-Ismail and Lin, 2012). Thus, in machine learning models, performing such validation methods with small subject pools runs the risk of overfitting and in turn may not be representative of the general population (De Boer *et al.*, 2021). The present study also included a relatively small subject pool, which was however greater than most psychosis studies assessing acoustic impairments. To overcome the issue of overfitting, we employed a bootstrapping approach and to examine the potential role of outliers in the model outcome (accuracy). Bootstrapping has been suggested as an appropriate procedure to test model performance in the presence of small samples since it reduces the chances of bias in model generalisation (Low, Bentley and Ghosh, 2020) and hence to evaluate whether model performance is due to a true group effect or whether it is due to bias in the data (Efron and Tibshirani, 1994). An additional reason for the use of this method emerged from the visual inspection of Q-Q plots which indicated that the current data was not normally distributed. Such non-linear distribution, hints at the possibility of spurious effects (presence of outliers); for this reason, bootstrapping is an ideal approach to examine the presence of bias in the data.

The modelling results indicate that the ability of acoustic features to identify CHR-P individuals is ill-defined and does not seem to emerge from our sample. Indeed, the CHR-N cohort was included to account for the presence of comorbidities in CHR-Ps. It was expected that the effects emerging from the CHR-P and CHR-N comparison would indicate the specificity of acoustic effects in CHR-Ps; instead, the present results remain puzzling and hint at a spurious effect of acoustic impairments in the CHR-N group. Indeed, evaluating model performance in clinical cohorts, which are highly heterogeneous, must take into consideration the individual influence of variability driven by the likely heterogeneous clinical cohort under study, potentially leading to group differences driven by disparate symptom patterns rather than true group effects.

4.5 Correlations.

Prior to multiple corrections, significant correlation results were obtained between prosodic values and the SPI-A, CAARMS total symptom severity as well as for the NBI, PA and DS CAARMS subscales.

Apart from the SPI-A total symptom severity item, the clinical features that showed significant associations with prosodic metrics, consisted of the CAARMS which assesses positive symptoms of psychosis (Yung *et al.*, 2005). This is opposed to the majority of studies that reported an association between acoustic features and negative symptoms such as blunted vocal affect or alogia in CHR-Ps (Stanislawski *et al.*, 2021) and in ScZ (Rapcan *et al.*, 2010; Covington *et al.*, 2012; Bernardini *et al.*, 2016). The bulk of the literature has largely observed a link between acoustic impairments and negative symptoms in psychosis, given that negative features are the constructs that temporal and prosodic indices are supposed to be measuring (Cohen et al., 2020). On the other hand, the relationship between positive clinical features and temporal and prosodic metrics has not been widely investigated given the loose conceptual link between them. Indeed, as suggested by Agurto *et al.* (2020), positive clinical symptoms would likely present stronger links with the content of speech rather than with temporal and prosodic features. Yet, one previous study in CHR-Ps revealed a link between the temporal measure of turn-taking and positive symptoms indicating higher positive symptomatology in individuals with aberrant turn-taking (Sichlinger *et al.*, 2019). Yet, it has been suggested that aberrant turn-taking and pausing behaviour is an acoustic feature not related to alogia but rather to positive symptoms of psychosis (Andreasen and Grove, 1986). The association between negative symptoms and prosodic measures found in the present manuscript indicates that prosodic impairments were observed in CHR-Ps with lower positive CAARMS symptom items, which suggests that acoustic impairments were not associated with higher clinical symptomatology in CHR-Ps, thus compromising the clinical utility of these correlations.

In addition, most clinical features that were associated to pitch percentile values indicated that higher symptomatology was associated with lower 5^{th} and 25^{th} pitch percentiles. However, these prosodic metrics suggest that participants with higher psychopathology presented lower pitch values. In this case, rather than indicating speech aberrations in CHR-Ps, lower pitch values could be explained by other non-language features, such as gender and age and may instead indicate that CHR-Ps with higher symptomatology had lower pitch because they consisted for the most part of male and/or older individuals. The onset of positive symptoms commonly occurs many years after the onset of negative symptoms (Yung *et al.*, 2019), therefore, it is plausible that older CHR-P individuals presented higher positive symptoms. However, it is generally assumed that positive sub-threshold symptoms are more frequent in female participants, while males often present more severe negative symptoms and social impairments (Beck *et al.*, 2019).

On the other hand, the DS CAARMS item correlated negatively with both 5th and 95th percentiles. This indicates greater pitch variability in CHR-P individuals with higher DS items scores (aka tangentiality, derailment, aberrant word usage) and is consistent with the conceptual association of positive symptoms like positive thought disorder to acoustic features such as increased/pressured speech and higher vocal variability (Hitczenko, Mittal and Goldrick, 2021).

4.5.1. Correlations following multiple comparison correction.

The correlations between clinical functional and acoustic metrics that initially reached significance did not survive multiple comparisons corrections. Previous ScZ studies are in accordance with the lack of robust associations between clinical features and acoustic metrics. Despite the wide number of studies reporting associations between symptomatology and acoustic features, a consistent share of studies has not been able to replicate such a link (Rapcan *et al.*, 2010; Covington *et al.*, 2012; Arevian *et al.*, 2020). Indeed, Cohen *et al.* (2016, 2020) attributed the lack of convergence between acoustic and clinical measures to the presence of "*different resolutions*" of these indices: while clinicians can potentially infer a holistic picture of the patient's symptomatology, the quantitative acoustic analysis examines a speech sample within a constrained time window without taking into account the contextual, environment as well as temporal nature of acoustic signatures (Cohen *et al.*, 2016).

Overall, given that CHR-P individuals are known to present more profound negative symptoms (Corcoran, Crusius and Mussweiler, 2011) and that the acoustic features examined tap into negative clinical features, the present correlation results may have shown more promising findings if the association of acoustic features was investigated in relation to negative rather than positive symptoms. This was however not possible in the present work, given that negative clinical features were not assessed in the sample.

4.6 Strengths.

4.6.1. CHR-Ns: an "active" control group.

One strength of this study consists in the examination of CHR-N as an "active" control group given that group effects emerging from the comparison between CHR-Ps and HCs may be driven by features unspecific to psychosis but driven by the high comorbidity present in the CHR-P population. Including such cohort was consistent with previous studies (Millman *et al.*, 2019; Haining, Brunner, *et al.*, 2021). To our knowledge, this is the first study within

the literature on speech disturbances in psychosis that compared psychosis individuals to a group presenting a range of psychopathologies to examine the specificity of the disorder in addition to the typical control group comparison. This addresses the sensitivity of automated acoustic analysis methods and their specificity to psychosis versus other illnesses (Hitczenko, Mittal and Goldrick, 2021). Despite the evidence of acoustic deficits in our CHR-Ns that the current studies seem to hint to, such effect was clarified by the null findings observed from the bootstrapping analysis. Future studies should further investigate the specificity of acoustic features in psychosis and psychosis risk cohorts.

4.6.2 Collinearity across predictors.

The iterative approach utilised in the present study for variable selection based on multicollinearity across predictors represents an additional strength to this thesis. Accounting for multicollinearity is useful in explanatory models given that collinearity across predictors can lead to inflated standard errors affecting the fit of the model. This is central in models prior to validation, whilst multicollinearity is not as damning for predictive modelling (Shmueli, 2010). Provided that a large part of the work on acoustic aberrations in psychosis has not performed any validation procedures (Compton *et al.*, 2018; Sichlinger *et al.*, 2019; Agurto *et al.*, 2020; de Boer, van Hoogdalem, *et al.*, 2020; de Boer, Voppel, *et al.*, 2020) and considering the abundance and plausible redundancy of acoustic features that have been fed into models to identify vocal signatures of psychosis (Hitczenko, Mittal and Goldrick, 2021), the issue of accounting for multicollinearity becomes crucial. Future studies should be mindful of the collinearity across predictions and how this could inflate the model performance. To address this, a method for variable selection, such as the one used in the present thesis, may be used in addition to validation approaches to assess the accuracy of the models built.

4.7 Limitations.

The present study's findings should be interpreted with caution in light of the following considerations.

Many individuals across groups consisted of female participants. The uneven gender composition could be due to a greater willingness to engage in help-seeking behaviours and being more prone to communicate mental health difficulties (Mackenzie, Gekoski and Knox, 2006). However, gender can have a considerable influence on the acoustic parameters (Titze,

1989) and although gender was considered as a covariate in the binary logistic regression, future studies should aim to obtain more balanced samples in terms of gender characteristics.

The present study is limited in the inability to use appropriate validation methods such as cross-validation. Bootstrapping was chosen given that cross-validation methods would likely produce biased estimates of model generalisation with small sample sizes (Low, Bentley and Ghosh, 2020). Therefore, our regression analyses may have benefitted from increasing the number of participants to determine the true predictability of our models. Yet, with small sample sizes, widely reported in the extant literature on acoustic markers of psychosis (sample size ranging between 20-30 participants; Parola *et al.*, 2020), bootstrapping is an appropriate method to estimate model performance.

Ideally, research would be conducted in much larger samples with the aim of collecting data for language analysis purposes. Automated acoustic analysis currently requires speech to be recorded under very good conditions and different recording conditions across studies can result in incompatible data (Hitczenko, Mittal and Goldrick, 2021). Careful cleaning of the data during the pre-processing stage of the present study aimed at reducing sources of noise and reverb that would have affected the results. Nevertheless, the original recordings presented a substantial amount of noise and reverb which might have still affected the results.

An additional limitation of the present study which may partially account for our results consists in the speech task employed. Eliciting speech from clinical interviews is a widely utilised method across the psychosis literature on speech impairments (Sichlinger *et al.*, 2019; Tahir *et al.*, 2019). Yet, speech elicitation methods of clinical assessments such as the CAARMS, specifically designed for the identification of individuals with psychotic psychopathology or at risk of developing such disorders (Yung *et al.*, 2005), might introduce variability across groups, given that the assessment is clearly targeted to CHR-Ps. It is plausible that different arousal levels (Cohen *et al.*, 2020) may emerge between CHR-Ps and the other two included groups since, during the CAARMS, participants who did meet the CHR-P criteria, found the questions asked to be relevant to their mental health difficulties and thus would likely present more discomfort in disclosing such personal/sensitive information compared to those who did not meet the criteria and did not have to disclose such personal information. For this reason, most studies eliciting speech from interviews have used emotionally neutral content to avoid the confounding factor that the nature of the interview can elicit across groups (Cohen *et al.*, 2020; De Boer *et al.*, 2021).

In relation to this, previous work has highlighted the importance of accounting for the context and content of speech elicitation methods, given that acoustic impairments become more pronounced in contexts with higher cognitive and social demand, but such effects disappear after accounting for contextual aspects (Cohen *et al.*, 2016). Indeed, in the present work, the familiarity of the content spoken might consist of an additional confound. It is likely that meeting the psychosis risk criteria, CHR-Ps found the questions asked more relevant to their own experiences and might therefore answer faster, with fewer pauses and with an enhanced speaking production. On the other hand, HC and CHR-N participants may have reported more hesitation pauses given that the questions they were asked referred to experiences they would be less "familiar" with or would not know how to answer. Future studies should therefore consider employing more controlled paradigms to elicit speech content (Cohen *et al.*, 2020).

Future work should be mindful of the acoustic effects derived from the speech elicitation methods. To achieve this, a range of different approaches can be used to remove the task-related confounds on vocal production as shown by Cohen's lab (2016; 2021), for instance, by examining a variety of tasks. Eliciting speech with narrative description as well as reading aloud can examine the presence of alogia versus a rather general lack of motivation and energy (Lysaker and Bell, 1995; Trémeau *et al.*, 2013), while examining sustained phonation as well as continuous speech can detect whether the presence of pitch instability (jitter, shimmer, etc.) emerges from aberrations in the motor control of the vocal fold. Therefore, using different tasks with disparate cognitive and social characteristics can allow for the assessment of contextual and mechanistic aspects of speech without being specifically bound to a single speech elicitation context and the associated confounding factors (Parola *et al.*, 2020).

4.8 Clinical implications.

One clinical implication of the present study consists in the lack of assessment of negative symptoms. This is particularly important to examine the mechanisms underlying acoustic impairments given that acoustic aberrations in CHR-Ps are conceptually related to negative symptoms of psychosis such as blunted vocal affect or alogia (Cohen *et al.*, 2020; Parola *et al.*, 2020). The lack of comparison of the acoustic features observed and negative symptoms hinders the correlational results' interpretability.

In addition, most CHR-P participants in the YouR-study were recruited from the community (McDonald *et al.*, 2019). The transition rate for the whole YouR-study sample was found to be at ≈7%, consisting of a significantly lower transition rate compared to previous findings

obtained from clinical samples (22%; Fusar-Poli *et al.*, 2012). The divergent transition rates raise the question of whether there is a difference in transition rates to psychosis between the community-recruited versus clinical-help seeking individuals. Accordingly, previous studies have shown that clinical versus community-recruited CHR-Ps may differ in transition rates (Fusar-Poli, Schultze-Lutter, *et al.*, 2016) and thus it remains plausible that the sample in the present study may be less psychosis enriched. However, CHR-P participants in the YouR-cohort are characterized by cognitive, clinical and physiological alterations that are consistent with previous findings from CHR-P cohorts recruited through the clinical pathways (Haining *et al.*, 2020; Grent *et al.*, 2021; Haining, Karagiorgou, *et al.*, 2021).

4.9 Future directions.

The findings of the present work may be informative for future research. Replications of our results should be conducted in larger samples. It would be useful to investigate acoustic abnormalities not only in comparison with HCs but also in CHR-P subgroups to observe how acoustic measures could relate to transition rates to determine clinical utility. Moreover, it may be useful to investigate acoustic aberrations in relation to the underlying physiological mechanisms. Further investigation should be conducted to examine whether atypical acoustic production is associated with neuromotor control, antipsychotic medication or auditory processing (Cannizzaro *et al.*, 2004; Konopka and Roberts, 2016; Corcoran *et al.*, 2020).

Furthermore, speech involves not only acoustic aspects but also lexical, semantic and syntactic aspects which have been widely shown to be aberrant in ScZ and CHR-Ps (i.e. tangentiality and poverty of content) (Bedi *et al.*, 2015; Mota, Copelli and Ribeiro, 2017; Corcoran *et al.*, 2018). Investigating acoustic as well as semantic and syntactic abnormalities may yield more promising results. In accordance, Çokal *et al.* (2019) observed that only syntactically motivated pauses were impaired in ScZ patients compared to their healthy counterparts. Further insight into the acoustic signatures of psychosis may be achieved in the investigation of a holistic examination of speech characteristics which may include semantic, lexical and acoustic aspects of language production (Parola *et al.*, 2020).

The lack of acoustic signatures of psychosis that emerges from the present study, highlights the limited understanding that there is on acoustic aberrations within the early stages of psychosis. Emerging from the present and previous work there has been an over-reliance on classification with the sole purpose to categorise individuals into two groups based on the

presence or absence of acoustic impairments (Hitczenko, Mittal and Goldrick, 2021). Clearly, acoustic and speech impairments are not dichotomous in nature and while some psychosis individuals do not present these aberrations, some healthy controls do (Andreasen and Grove, 1986). Therefore, there is a need to shift towards the examination of acoustic features by focusing on the comparison with clinical, functional and neurocognitive features and investigating how specific these measures are to psychosis versus other illnesses as well as understanding their predictive value (Low, Bentley and Ghosh, 2020).

4.10 Conclusion.

The present study builds on the emerging bulk of literature on acoustic aberrations in psychosis. Our finding suggests that temporal and prosodic aspects of speech are not impaired in CHR-P participants. Additionally, the acoustic features examined between CHR-P and CHR-Ns indicated the presence of acoustic impairments in CHR-Ns. Initially, the regression models based on prosodic and especially on temporal features obtained acceptable discriminant ability (accuracy). Yet, employing bootstrapping, all the acoustic models failed to maintain significant above-chance diagnostic accuracy. The spurious classification accuracy of the original models indicates the importance of employing validation methods especially given the novel and still limited evidence on acoustic signatures of psychosis (Agurto *et al.*, 2020; Corcoran *et al.*, 2020). Due to the prevalence of small sample sizes across the literature and heterogeneity of the clinical groups examined, it is advised to employ bootstrapping. However, given that cross-validation methods are needed for true predictability, larger groups are required to truly test acoustic features as biomarkers of early psychosis.

Overall, given the absence of acoustic signatures of at-risk psychosis found in the present work, and the clearer evidence of semantic/syntactic impairments in CHR-Ps (Elvevåg *et al.*, 2007; Gupta *et al.*, 2018), it may be speculated that the semantic and syntactic features may constitute a more promising biomarker of early psychosis. Future studies should clarify whether acoustic abnormalities are present in sub-groups of CHR-P participants with elevated psychosis-risk.

Appendices

Appendix A

Figure A1-A4 depict Q-Q plots for temporal and prosodic features before and after correction.

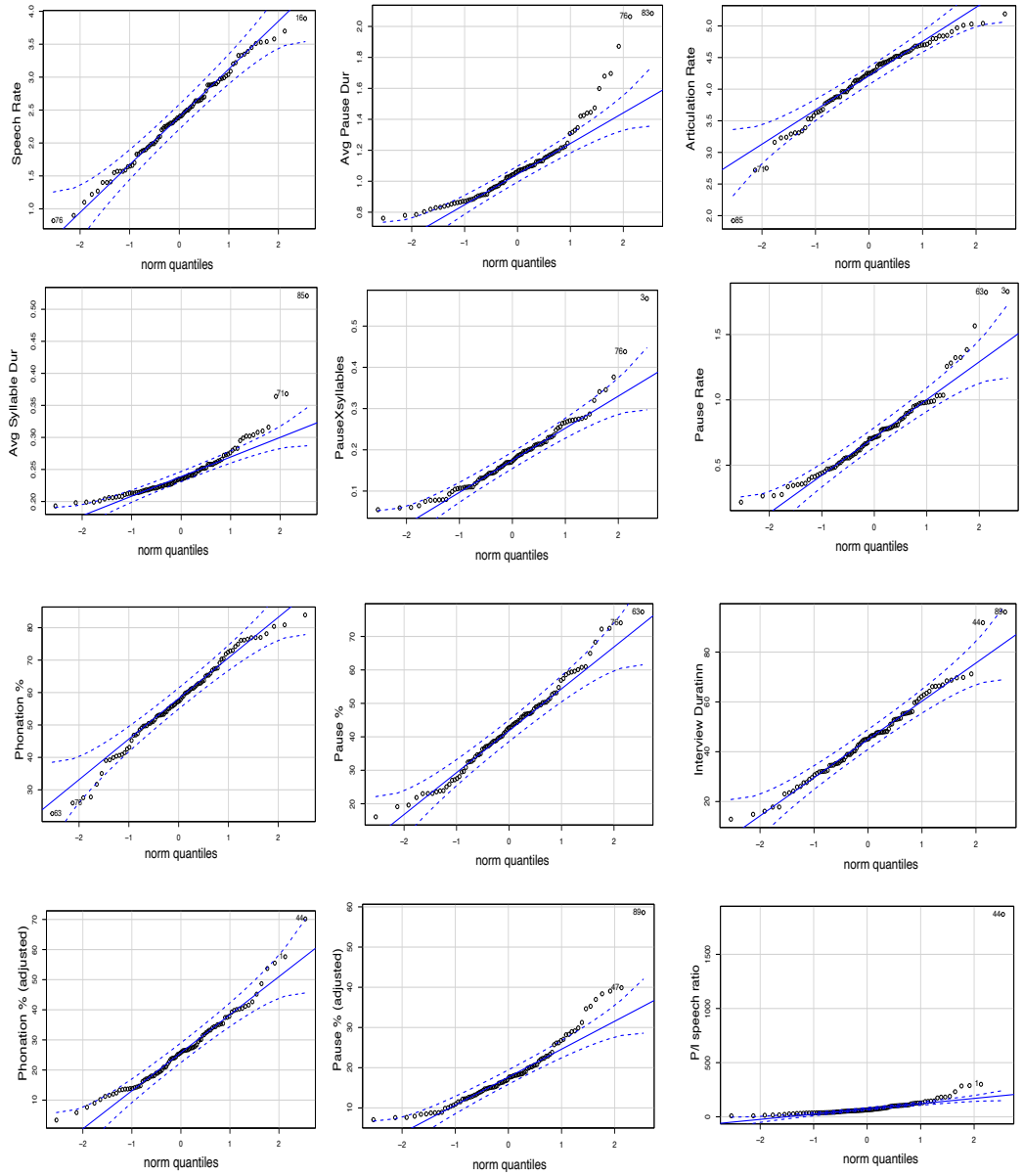

Figure A 1 Q-Q plots of temporal features before correction. *Note:* P/I, participant/interviewer; %, percentage; Avg, average.

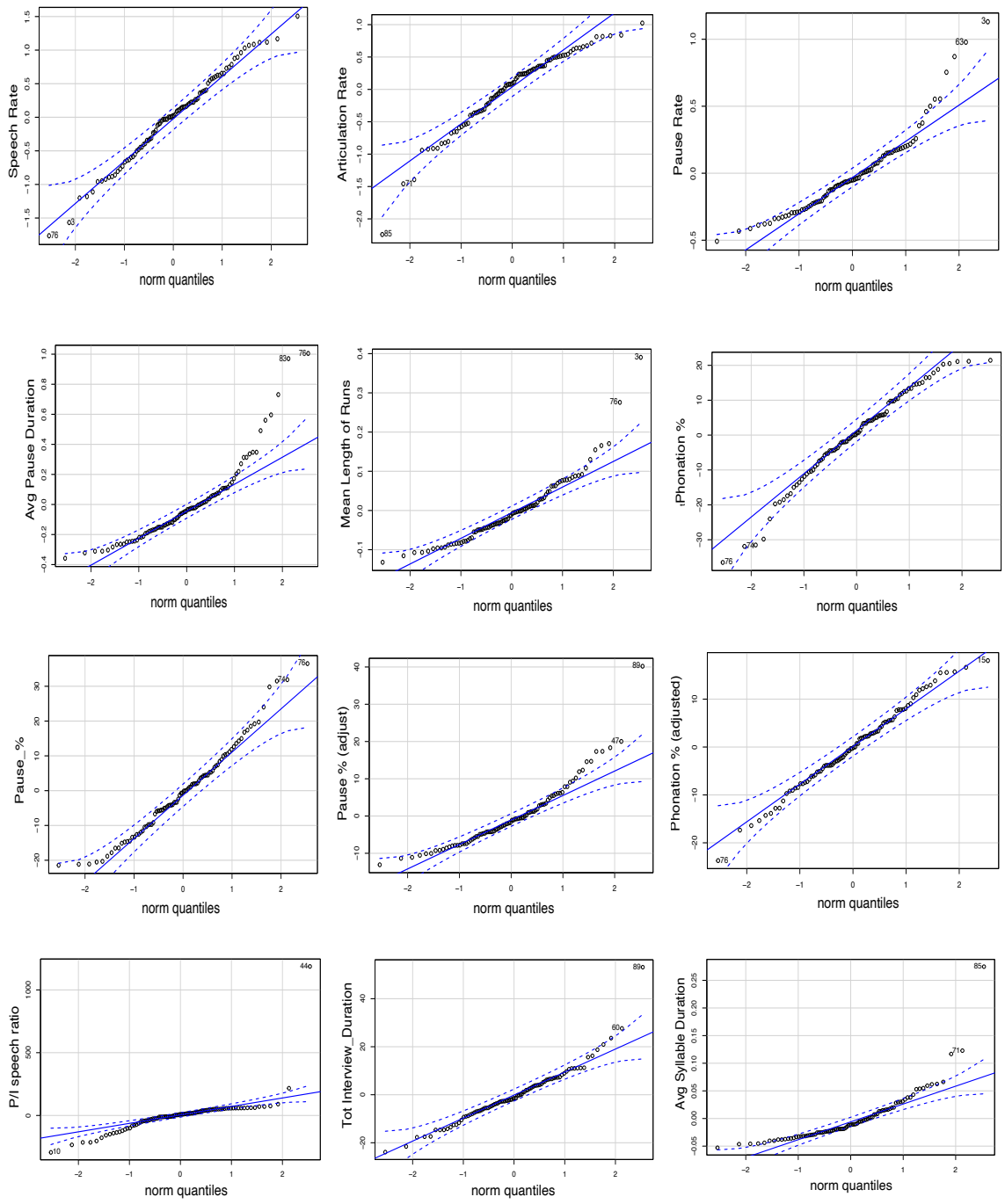

Figure A 2 Q-Q plots of temporal features after correction. *Note:* P/I, participant/interviewer; %, percentage; Avg, average.

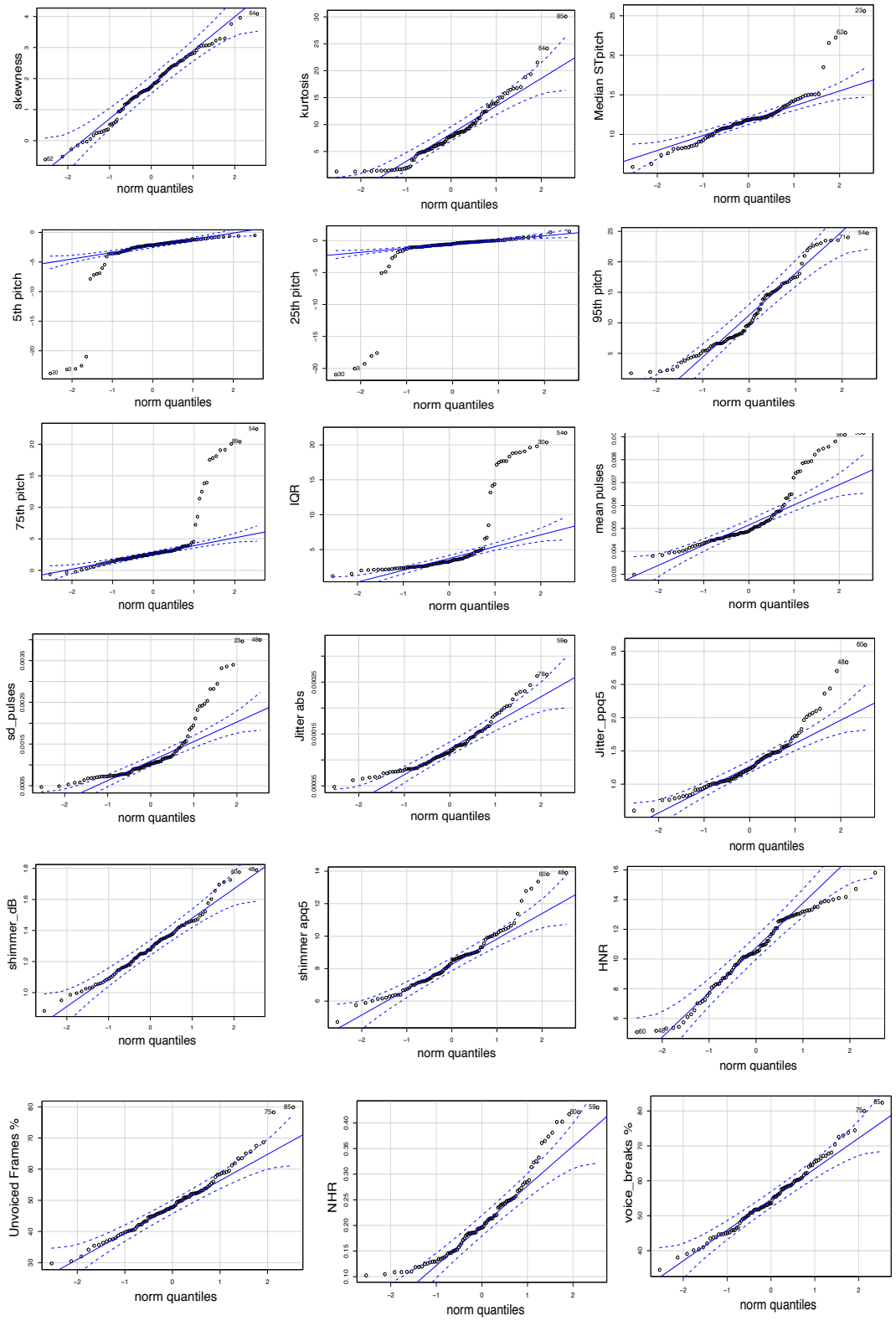

Figure A 3 Q-Q plots of prosodic features before correction. *Note*%, percentage; dB, decibels; sd; standard deviation; ppq5, five-point Period Perturbation Quotient; apq5, ; five-point Amplitude Perturbation Quotien; ST, semitones; IQR, interquartile range; abs, absolute; NHR, noise to harmonics ratio; HNR, harmonics to noise ratio.

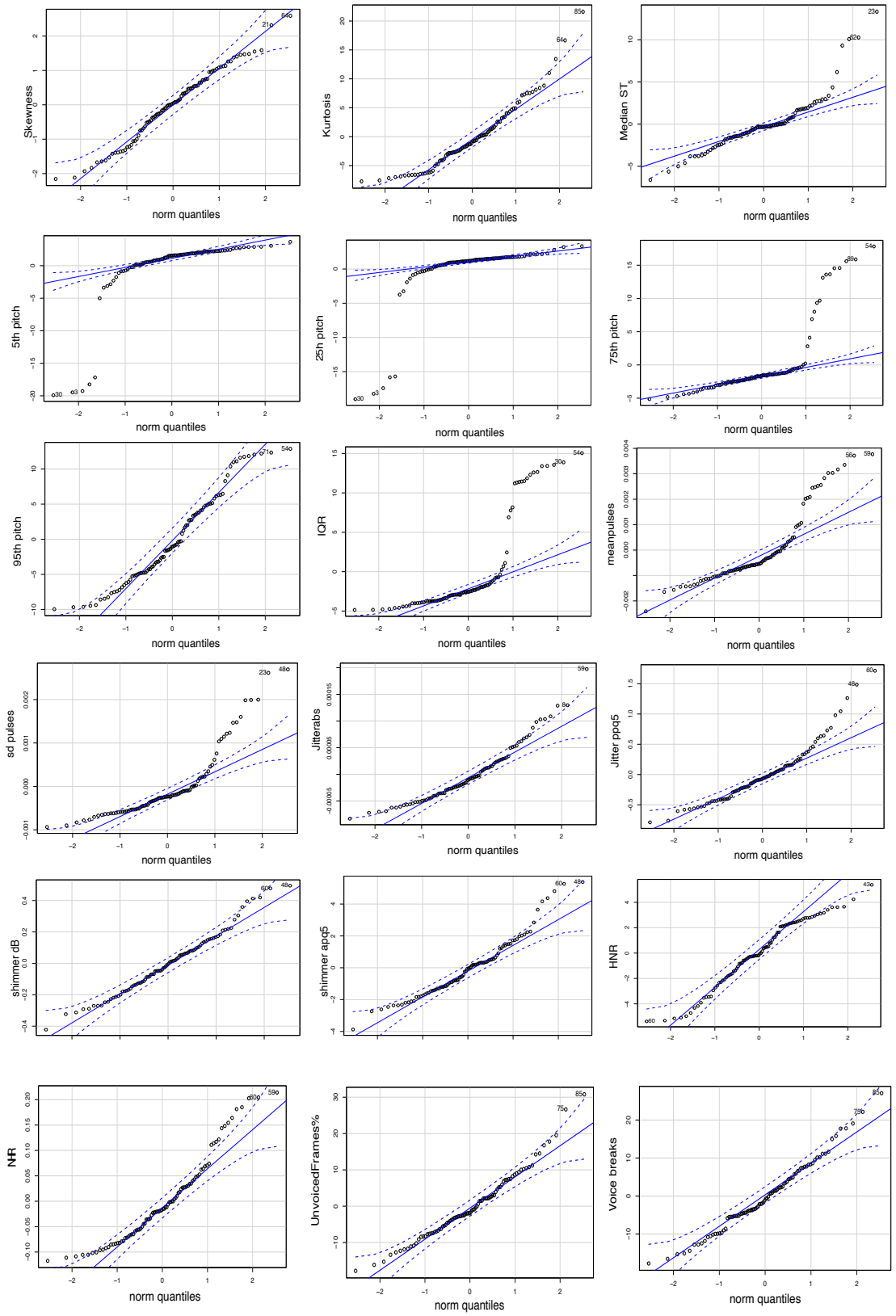

Figure A 4 Q-Q plots of prosodic features after correction. *Note*%, percentage; dB, decibels; sd; standard deviation; ppq5, five-point Period Perturbation Quotient; apq5, ; five-point Amplitude Perturbation Quotien; ST, semitones; IQR, interquartile range; abs, absolute; NHR, noise to harmonics ratio; HNR, harmonics to noise ratio.

Appendix Table B.1

P-values of Kendall's tau correlations between all acoustic features and severity for each CAARMS subscale item

Acoustic feature	UTC	NBI	PA	DS
Mean pulses	0.85782814	0.70280412	0.066069025	0.02764441
SD pulses	0.74314284	0.16531656	0.984820005	0.26901452
Jitter absolute	0.82366618	0.49881038	0.288382838	0.60364050
Jitter ppq5	0.47095419	0.72870200	0.515237198	0.30302356
Shimmer absolute	0.89917116	0.71571217	0.942362759	0.52998352
Shimmer apq5	0.92689318	0.75491330	0.900068693	0.79518297
NHR	0.91995301	0.75491330	0.486198353	0.83417992
HNR	0.54363412	0.55542562	0.335678835	0.58623940
Unvoiced local (%)	0.78311405	0.59083828	0.347265455	0.36580195
Voice breaks (%)	0.60920888	0.60286912	0.400365250	0.48704204
Skewness Pitch	0.26897678	0.46637450	0.216191328	0.11350387
Kurtosis Pitch	0.50940927	0.55542562	0.555311365	0.03556984
25th Pitch	0.16075660	0.03742328	0.634315731	0.13174023
5th pitch	0.38937927	0.03151674	0.328094339	0.02225230
Median Pitch (ST)	0.70385237	0.62725523	0.383529557	0.19144891
95th Pitch	0.87157163	0.87596704	0.586333803	0.02591940
IQR Pitch	0.44972028	0.27457564	0.495778536	0.10426759
75th Pitch	0.13396313	0.50988287	0.043320257	0.04358361
Speech duration (%)	0.17419765	0.08921031	0.047417036	0.59781436
Avg Syllable Duration	0.77640908	0.70280412	0.852084663	0.46629768
Articulation Rate	0.83047408	0.75491330	0.900068693	0.50292160
Speech Rate	0.22283559	0.35801210	0.047417036	0.86699633
Avg Pause Duration	0.02081425	0.71571217	0.008745524	0.84727520
Mean length of runs	0.38937927	0.17066565	0.158019876	0.97995825
Pause Rate	0.42903999	0.11059548	0.174308718	0.42146634

Pause Duration (%)	0.17419765	0.08921031	0.047417036	0.59781436
Tot interview Duration (%)	0.29234516	0.79475616	0.540106671	0.41185854
Pause duration (adjusted; %)	0.11674975	0.35801210	0.019270662	0.87358912
Speech Duration (adjusted; %)	0.80332393	0.34016217	0.540106671	0.35699176
P/I speech ratio	0.03055794	0.54385609	0.043320257	0.04813159

Appendix Table B.2
P-values of Kendall's tau correlations between all acoustic features and clinical and functional measures

Acoustic feature	CAARMS	SPI-A	GAF	GF: Role	GF: Social
Mean pulses	0.338	0.338	0.61937001	0.828291719	0.6829960074
SD pulses	0.599	0.460	0.43922627	0.124719400	0.1993259005
Jitter absolute	0.463	0.502	0.31360836	0.369536610	0.5049382359
Jitter ppq5	0.623	0.313585277	0.10414965	0.520760572	0.5995430560
Shimmer absolute	0.938	0.449771719	0.79288636	0.630824357	0.9867010774
Shimmer apq5	0.938	0.420137894	0.97169566	0.537461240	0.8090178655
NHR	0.711	0.579407661	0.76566802	0.456730840	0.8479868064
HNR	0.517	0.602604179	0.57991144	0.642966124	0.5484622620
Unvoiced local (%)	0.816	0.134967727	0.54167319	0.274406665	0.3906563233
Voice breaks (%)	0.924	0.201794762	0.81484599	0.436413876	0.4582398928
Skewness Pitch	0.181	0.290003507	0.82035944	0.039149690	0.6526747579
Kurtosis Pitch	0.432	0.545420530	0.64461084	0.059566018	0.6170363607
25th Pitch	0.007	0.009233215	0.57506315	0.624790376	0.6347497413
5th pitch	0.0032	0.011209013	0.48677912	0.935598399	0.9933503078
Median Pitch (ST)	0.347	0.590953574	0.59949370	0.092985008	0.5049382359
95th Pitch	0.635	1.000000000	0.13430831	0.285766932	0.0705237101
IQR Pitch	0.422	0.099773025	0.21428877	0.017845323	0.0552535030
75th Pitch	0.356	0.959814492	0.06199438	0.029132523	0.0022858593

Speech duration (%)	0.993	0.626209920	0.34165067	0.493532310	0.2605366119
Avg Syllable Duration	0.698	0.382464775	0.50928008	0.347284378	0.3906563233
Articulation Rate	0.776	0.329988705	0.53231409	0.374091175	0.4046050933
Speech Rate	0.672	0.480555569	0.15999921	0.285766932	0.0955442445
Avg Pause Duration	0.575	0.373380337	0.72805019	0.156721691	0.3420591705
Mean length of runs	0.993	0.674568386	0.06818868	0.548742582	0.2399430233
Pause Rate	0.924	0.840272048	0.20399778	0.667535215	0.2975131227
Pause Duration (%)	0.993	0.626209920	0.34165067	0.493532310	0.2605366119
Tot interview Duration (%)	0.763	0.355616362	0.41852543	0.976250780	0.7897032397
Pause duration (adjusted; %)	0.85	0.470168608	0.41852543	0.606840032	0.3378415768
Speech Duration (adjusted; %)	0.856	0.686881267	0.93213736	0.874966055	0.3505949138
P/I speech ratio	0.412	0.173693198	0.00589838	0.007664372	0.0009097939

References

Abu-Mostafa, Y. S., Magdon-Ismail, M. and Lin, H.-T. (2012) *Learning from data*. AMLBook New York.

Agnew-Blais, J. and Seidman, L. J. (2013) 'Neurocognition in youth and young adults under age 30 at familial risk for schizophrenia: a quantitative and qualitative review', *Cognitive neuropsychiatry*, 18(1–2), pp. 44–82.

Agurto, C. *et al.* (2020) 'Analyzing acoustic and prosodic fluctuations in free speech to predict psychosis onset in high-risk youths', in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. doi: 10.1109/EMBC44109.2020.9176841.

Akbarian, S. *et al.* (1996) 'Selective alterations in gene expression for NMDA receptor subunits in prefrontal cortex of schizophrenics', *Journal of Neuroscience*, 16(1), pp. 19–30.

Alghowinem, S. *et al.* (2013) 'Detecting depression: a comparison between spontaneous and read speech', in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 7547–7551.

Allen, R. M. and Young, S. J. (1978) 'Phencyclidine-induced psychosis.', *The American journal of psychiatry*.

Alpert, M. *et al.* (2000) 'Prosody and lexical accuracy in flat affect schizophrenia', *Psychiatry research*, 97(2–3), pp. 107–118.

Alpert, M., Kotsaftis, A. and Pouget, E. R. (1997) 'Speech fluency and schizophrenic negative signs', *Schizophrenia Bulletin*, 23(2), pp. 171–177.

Andreasen, N. C. (1979) 'Thought, language, and communication disorders: I. Clinical assessment, definition of terms, and evaluation of their reliability', *Archives of general Psychiatry*, 36(12), pp. 1315–1321.

Andreasen, N. C. (1984) 'Scale for the assessment of positive symptoms', *Group*, 17(2), pp. 173–180.

Andreasen, N. C. (1989) 'The Scale for the Assessment of Negative Symptoms (SANS): conceptual and theoretical foundations', *The British journal of psychiatry*, 155(S7), pp. 49–52.

Andreasen, N. C. and Grove, W. M. (1986) 'Thought, language, and communication in

schizophrenia: diagnosis and prognosis', *Schizophrenia bulletin*, 12(3), pp. 348–359.

Appelbaum, P. S., Robbins, P. C. and Roth, L. H. (1999) 'Dimensional approach to delusions: comparison across types and diagnoses', *American Journal of Psychiatry*, 156(12), pp. 1938–1943.

Arevian, A. C. *et al.* (2020) 'Clinical state tracking in serious mental illness through computational analysis of speech', *PloS one*, 15(1), p. e0225695.

Association, A. P. (2013) 'Desk reference to the diagnostic criteria from DSM-5'.

Avramopoulos, D. (2018) 'Recent advances in the genetics of schizophrenia', *Complex Psychiatry*, 4(1), pp. 35–51.

Bacon, S. F., Collins, M. J. and Plake, E. V (2002) 'Does the Global Assessment of Functioning assess functioning?', *Journal of Mental Health Counseling*, 24(3).

Balu, D. T. (2016) 'The NMDA receptor and schizophrenia: from pathophysiology to treatment', *Advances in pharmacology*, 76, pp. 351–382.

Barch, D. M. and Berenbaum, H. (1997) 'The effect of language production manipulations on negative thought disorder and discourse coherence disturbances in schizophrenia', *Psychiatry Research*, 71(2), pp. 115–127.

Bauer, S. M. *et al.* (2011) 'Culture and the prevalence of hallucinations in schizophrenia', *Comprehensive psychiatry*, 52(3), pp. 319–325.

Bearden, C. E. *et al.* (2011) 'Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis', *Journal of the American Academy of Child & Adolescent Psychiatry*, 50(7), pp. 669–680.

Beck, K. *et al.* (2019) 'Clinical and functional long-term outcome of patients at clinical high risk (CHR) for psychosis without transition to psychosis: A systematic review', *Schizophrenia research*, 210, pp. 39–47.

Bedi, G. *et al.* (2015) 'Automated analysis of free speech predicts psychosis onset in high-risk youths', *npj Schizophrenia*, 1(1), pp. 1–7.

Ben-Zeev, D. and Atkins, D. C. (2017) 'Bringing digital mental health to where it is needed most', *Nature human behaviour*, 1(12), pp. 849–851.

Bentall, R. P. *et al.* (2001) 'Persecutory delusions: a review and theoretical integration', *Clinical psychology review*, 21(8), pp. 1143–1192.

Bernardini, F. *et al.* (2016) 'Associations of acoustically measured tongue/jaw movements and portion of time speaking with negative symptom severity in patients with schizophrenia in Italy and the United States', *Psychiatry research*, 239, pp. 253–258.

Bleuler, E. (1911) *Dementia praecox, oder Gruppe der Schizophrenien*. Deuticke.

de Boer, J. N., Voppel, A. E., *et al.* (2020) 'Language disturbances in schizophrenia: the relation with antipsychotic medication', *npj Schizophrenia*, 6(1), pp. 1–9. doi: 10.1038/s41537-020-00114-3.

de Boer, J. N., van Hoogdalem, M., *et al.* (2020) 'Language in schizophrenia: relation with diagnosis, symptomatology and white matter tracts', *npj Schizophrenia*, 6(1), pp. 1–10. doi: 10.1038/s41537-020-0099-3.

De Boer, J. N. *et al.* (2021) 'Acoustic speech markers for schizophrenia-spectrum disorders: A diagnostic and symptom-recognition tool', *Psychological Medicine*. doi: 10.1017/S0033291721002804.

Boersma, P. (1993) 'Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound', in *Proceedings of the institute of phonetic sciences*. Citeseer, pp. 97–110.

Bora, E. and Murray, R. M. (2014) 'Meta-analysis of cognitive deficits in ultra-high risk to psychosis and first-episode psychosis: do the cognitive deficits progress over, or after, the onset of psychosis?', *Schizophrenia bulletin*, 40(4), pp. 744–755.

Bowie, C. R. and Harvey, P. D. (2006) 'Administration and interpretation of the Trail Making Test', *Nature protocols*, 1(5), pp. 2277–2281.

Bromberg-Martin, E. S., Matsumoto, M. and Hikosaka, O. (2010) 'Dopamine in motivational control: rewarding, aversive, and alerting', *Neuron*, 68(5), pp. 815–834.

Cannizzaro, M. *et al.* (2004) 'Voice acoustical measurement of the severity of major depression', *Brain and cognition*, 56(1), pp. 30–35.

Cannon, M., Jones, P. B. and Murray, R. M. (2002) 'Obstetric complications and schizophrenia: historical and meta-analytic review', *American Journal of Psychiatry*, 159(7), pp. 1080–1092.

Cannon, T. D. *et al.* (2002) 'Cortex mapping reveals regionally specific patterns of genetic and disease-specific gray-matter deficits in twins discordant for schizophrenia', *Proceedings of the National Academy of Sciences*, 99(5), pp. 3228–3233.

Caplan, R. *et al.* (1989) 'The kiddie formal thought disorder rating scale: clinical assessment, reliability, and validity', *Journal of the American Academy of Child & Adolescent Psychiatry*, 28(3), pp. 408–416.

Cardno, A. G. and Gottesman, I. I. (2000) 'Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics', *American journal of medical genetics*, 97(1), pp. 12–17.

Carlen, M. *et al.* (2012) 'A critical role for NMDA receptors in parvalbumin interneurons for gamma rhythm induction and behavior', *Molecular psychiatry*, 17(5), pp. 537–548.

Carrión, R. E. *et al.* (2011) 'Impact of neurocognition on social and role functioning in individuals at clinical high risk for psychosis', *American Journal of Psychiatry*, 168(8), pp. 806–813.

Carrión, R. E. *et al.* (2019) 'The global functioning: social and role scales—further validation in a large sample of adolescents and young adults at clinical high risk for psychosis', *Schizophrenia bulletin*, 45(4), pp. 763–772.

Castle, S. C. (2000) 'Clinical relevance of age-related immune dysfunction', *Clinical Infectious Diseases*, 31(2), pp. 578–585.

Catts, H. W. *et al.* (2016) 'Early identification of reading comprehension difficulties', *Journal of learning disabilities*, 49(5), pp. 451–465.

Cavelti, M. *et al.* (2018) 'Formal thought disorder is related to aberrations in language-related white matter tracts in patients with schizophrenia', *Psychiatry Research: Neuroimaging*, 279, pp. 40–50.

Chaika, E. O. (1990) *Understanding psychotic speech: Beyond Freud and Chomsky.* Charles C Thomas, Publisher.

Chang, M. *et al.* (2016) 'Voxel-based morphometry in individuals at genetic high risk for schizophrenia and patients with schizophrenia during their first episode of psychosis', *PLoS One*, 11(10), p. e0163749.

Chomsky, N. (1995) 'Language and nature', *Mind*, 104(413), pp. 1–61.

Cleynen, I. *et al.* (2021) 'Genetic contributors to risk of schizophrenia in the presence of a 22q11. 2 deletion', *Molecular psychiatry*, 26(8), pp. 4496–4510.

Cohen, A. S. *et al.* (2007) 'JOURNAL OF PSYCHIATRIC RESEARCH'.

Cohen, A. S. *et al.* (2015) 'Vocal acoustic analysis as a biometric indicator of information

processing: Implications for neurological and psychiatric disorders', *Psychiatry research*, 226(1), pp. 235–241.

Cohen, A. S. *et al.* (2016) 'Vocal expression in schizophrenia: Less than meets the ear.', *Journal of abnormal psychology*, 125(2), p. 299.

Cohen, A. S. *et al.* (2019) 'Ambulatory vocal acoustics, temporal dynamics, and serious mental illness.', *Journal of abnormal psychology*, 128(2), p. 97.

Cohen, A. S. *et al.* (2020) 'Using machine learning of computerized vocal expression to measure blunted vocal affect and alogia', *npj Schizophrenia*, 6(1), pp. 1–9. doi: 10.1038/s41537-020-00115-2.

Cohen, A. S., Kim, Y. and Najolia, G. M. (2013) 'Psychiatric symptom versus neurocognitive correlates of diminished expressivity in schizophrenia and mood disorders', *Schizophrenia research*, 146(1–3), pp. 249–253.

Cohen, A. S., Mitchell, K. R. and Elvevåg, B. (2014) 'What do we really know about blunted vocal affect and alogia? A meta-analysis of objective assessments', *Schizophrenia research*, 159(2–3), pp. 533–538.

Cohen, R. J. and Benedek, G. B. (1979) 'The functional relationship between the polymerization and catalytic activity of beef liver glutamate dehydrogenase: III. Analysis of Thusius' critique', *Journal of molecular biology*, 129(1), pp. 37–44.

Coid, J. W. *et al.* (2013) 'The relationship between delusions and violence: findings from the East London first episode psychosis study', *JAMA psychiatry*, 70(5), pp. 465–471.

Çokal, D. *et al.* (2019) 'Disturbing the rhythm of thought: Speech pausing patterns in schizophrenia, with and without formal thought disorder', *PLoS One*, 14(5), p. e0217404.

Compton, M. T. *et al.* (2018) 'The aprosody of schizophrenia: Computationally derived acoustic phonetic underpinnings of monotone speech', *Schizophrenia research*, 197, pp. 392–399.

Corcoran, C. M. *et al.* (2018) 'Prediction of psychosis across protocols and risk cohorts using automated language analysis', *World Psychiatry*, 17(1), pp. 67–75. doi: 10.1002/wps.20491.

Corcoran, C. M. *et al.* (2020) 'Language as a biomarker for psychosis: A natural language processing approach', *Schizophrenia Research*, 226(xxxx), pp. 158–166. doi: 10.1016/j.schres.2020.04.032.

Corcoran, K., Crusius, J. and Mussweiler, T. (2011) 'Social comparison: motives, standards, and mechanisms.'

Cornblatt, B. A. *et al.* (2007) 'Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia', *Schizophrenia bulletin*, 33(3), pp. 688–702.

Corretge, R. (2012) 'Praat vocal toolkit', *Barcelona, Spain: Praat. Retrieved from http://praatvocaltoolkit. com*.

Covington, M. A. *et al.* (2005) 'Schizophrenia and the structure of language: the linguist's view', *Schizophrenia research*, 77(1), pp. 85–98.

Covington, M. A. *et al.* (2012) 'Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders', *Schizophrenia research*, 142(1–3), pp. 93–95.

Cummings, A. and Čeponienė, R. (2010) 'Verbal and nonverbal semantic processing in children with developmental language impairment', *Neuropsychologia*, 48(1), pp. 77–85.

Cummins, N. *et al.* (2015) 'A review of depression and suicide risk assessment using speech analysis', *Speech communication*, 71, pp. 10–49.

Cuthbert, B. N. and Insel, T. R. (2013) 'Toward the future of psychiatric diagnosis: the seven pillars of RDoC', *BMC medicine*, 11(1), pp. 1–8.

Davies, G. *et al.* (2003) 'A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia', *Schizophrenia bulletin*, 29(3), pp. 587–593.

Davis, J. *et al.* (2016) 'A review of vulnerability and risks for schizophrenia: Beyond the two hit hypothesis', *Neuroscience & Biobehavioral Reviews*, 65, pp. 185–194.

Davison, A. C. and Hinkley, D. V. (1997) *Bootstrap methods and their application*. Cambridge university press.

Dean, K. and Murray, R. M. (2022) 'Environmental risk factors for psychosis', *Dialogues in clinical neuroscience*.

DeLisi, L. E. (2001) 'Speech disorder in schizophrenia: review of the literature and exploration of its relation to the uniquely human capacity for language', *Schizophrenia bulletin*, 27(3), pp. 481–496.

Denes, P. B. (1963) 'On the statistics of spoken English', *The Journal of the Acoustical Society of America*, 35(6), pp. 892–904.

Diagnostic and statistical manual of mental disorders : DSM-IV (no date). Fourth edition. Washington, DC : American Psychiatric Association, [1994] ©1994. Available at: https://search.library.wisc.edu/catalog/999733358502121.

Docherty, N. M. *et al.* (2003) 'Stability of formal thought disorder and referential communication disturbances in schizophrenia.', *Journal of Abnormal Psychology*, 112(3), p. 469.

Donoghue, J. F., Golowich, E. and Holstein, B. R. (2014) *Dynamics of the standard model*. Cambridge university press.

Dunn, O. J. (1964) 'Multiple comparisons using rank sums', *Technometrics*, 6(3), pp. 241–252.

Ebenezer, I. (2015) *Neuropsychopharmacology and therapeutics*. John Wiley & Sons.

Efron, B. and Tibshirani, R. J. (1994) *An introduction to the bootstrap*. CRC press.

Egerton, A. *et al.* (2017) 'Neuroimaging studies of GABA in schizophrenia: a systematic review with meta-analysis', *Translational psychiatry*, 7(6), pp. e1147–e1147.

Elvevåg, B. *et al.* (2007) 'Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia', *Schizophrenia research*, 93(1–3), pp. 304–316.

Eranti, S. V *et al.* (2013) 'Gender difference in age at onset of schizophrenia: a meta-analysis', *Psychological medicine*, 43(1), pp. 155–167.

Eyben, F. (2016) 'Acoustic features and modelling', in *Real-time Speech and Music Classification by Large Audio Feature Space Extraction*. Springer, pp. 9–122.

Fabiana, R. *et al.* (2021) 'Anodal Transcranial Direct Current Stimulation over the Cerebellum Enhances Sadness Recognition in Parkinson's Disease Patients: a Pilot Study'.

Farooq, S. *et al.* (2009) 'The relationship between the duration of untreated psychosis and outcome in low-and-middle income countries: a systematic review and meta analysis', *Schizophrenia research*, 109(1–3), pp. 15–23.

Foussias, G. *et al.* (2014) 'Negative symptoms of schizophrenia: clinical features, relevance to real world functioning and specificity versus other CNS disorders', *European Neuropsychopharmacology*, 24(5), pp. 693–709.

Fuller, B. F., Horii, Y. and Conner, D. A. (1992) 'Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety', *Research in nursing & health*, 15(5), pp. 379–389.

Fusar-Poli, P. *et al.* (2012) 'Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk', *Archives of general psychiatry*, 69(3), pp. 220–229.

Fusar-Poli, P. *et al.* (2013) 'The psychosis high-risk state: a comprehensive state-of-the-art review', *JAMA psychiatry*, 70(1), pp. 107–120.

Fusar-Poli, P., Cappucciati, M., *et al.* (2016) 'Heterogeneity of psychosis risk within individuals at clinical high risk: a meta-analytical stratification', *JAMA psychiatry*, 73(2), pp. 113–120.

Fusar-Poli, P., Schultze-Lutter, F., *et al.* (2016) 'The dark side of the moon: meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis', *Schizophrenia bulletin*, 42(3), pp. 732–743.

Fusar-Poli, P. *et al.* (2018) 'The science of prognosis in psychiatry: a review', *JAMA psychiatry*, 75(12), pp. 1289–1297.

Fusar-Poli, P. *et al.* (2020) 'Prevention of psychosis: advances in detection, prognosis, and intervention', *Jama Psychiatry*, 77(7), pp. 755–765.

Fusar-Poli, P. and Politi, P. (2008) 'Paul Eugen Bleuler and the birth of schizophrenia (1908)', *American Journal of Psychiatry*, 165(11), p. 1407.

Fusar-Poli, P. *et al.* (2015) 'At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction', *World Psychiatry*, 14(3), pp. 322–332.

Fusaroli, R. *et al.* (2017) 'Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis', *Autism Research*, 10(3), pp. 384–407.

Fusaroli, R. *et al.* (2019) 'Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development', *Cognition*, 183, pp. 1–18.

Gonzalez-Burgos, G., Hashimoto, T. and Lewis, D. A. (2010) 'Alterations of cortical GABA neurons and network oscillations in schizophrenia', *Current psychiatry reports*, 12(4), pp. 335–344.

Gonzalez-Burgos, G. and Lewis, D. A. (2012) 'NMDA receptor hypofunction, parvalbumin-positive neurons, and cortical gamma oscillations in schizophrenia', *Schizophrenia bulletin*, 38(5), pp. 950–957.

Green, M. F., Horan, W. P. and Lee, J. (2015) 'Social cognition in schizophrenia', *Nature Reviews Neuroscience*, 16(10), pp. 620–631.

Grent, T. *et al.* (2021) '40-Hz Auditory Steady-State Responses Characterize Circuit Dysfunctions and Predict Clinical Outcomes in Clinical-High-Risk Participants: A MEG Study', *Biological Psychiatry*.

Gross, G. (1987) *Bonn scale for the assessment of basic symptoms: BSABS; Manual, Kommentar, Dokumentationsbogen*. Springer.

Group, F.-N. B. W. (2016) 'BEST (Biomarkers, endpoints, and other tools) resource [Internet]'.

Grzenda, A. and Widge, A. S. (2020) 'Electroencephalographic biomarkers for predicting antidepressant response: new methods, old questions', *JAMA psychiatry*, 77(4), pp. 347–348.

Gupta, T. *et al.* (2018) 'Automated analysis of written narratives reveals abnormalities in referential cohesion in youth at ultra high risk for psychosis', *Schizophrenia Research*, 192, pp. 82–88. doi: 10.1016/j.schres.2017.04.025.

Häfner, H. (2003) 'Gender differences in schizophrenia', *Psychoneuroendocrinology*, 28, pp. 17–54.

Haining, K. *et al.* (2020) 'Neuropsychological deficits in participants at clinical high risk for psychosis recruited from the community: Relationships to functioning and clinical symptoms', *Psychological medicine*, 50(1), pp. 77–85.

Haining, K., Karagiorgou, O., *et al.* (2021) 'Prevalence and predictors of suicidality and non-suicidal self-harm among individuals at clinical high-risk for psychosis: Results from a community-recruited sample', *Early intervention in psychiatry*, 15(5), pp. 1256–1265.

Haining, K., Brunner, G., *et al.* (2021) 'The relationship between cognitive deficits and impaired short-term functional outcome in clinical high-risk for psychosis participants: A machine learning and modelling approach', *Schizophrenia Research*, 231, pp. 24–31. doi: 10.1016/j.schres.2021.02.019.

Hartsuiker, R. J., Pickering, M. J. and Veltkamp, E. (2004) 'Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals', *Psychological science*, 15(6), pp. 409–414.

Harvey, P. D. *et al.* (2006) 'Negative symptoms and cognitive deficits: what is the nature of their relationship?', *Schizophrenia bulletin*, 32(2), pp. 250–258.

Hasbi, A., O'Dowd, B. F. and George, S. R. (2011) 'Dopamine D1-D2 receptor heteromer

signaling pathway in the brain: emerging physiological relevance', *Molecular brain*, 4(1), pp. 1–6.

Hashimoto, K., Murata, K. and Yoshii, R. (2017) 'Out-of-time-order correlators in quantum mechanics', *Journal of High Energy Physics*, 2017(10), pp. 1–31.

Henriksen, M. G., Nordgaard, J. and Jansson, L. B. (2017) 'Genetics of schizophrenia: overview of methods, findings and limitations', *Frontiers in human neuroscience*, 11, p. 322.

Hinney, B. *et al.* (2021) 'Does hippocampal volume predict transition to psychosis in a high-risk group? A meta-analysis', *Frontiers in Psychiatry*, 11, p. 614659.

Hitczenko, K., Mittal, V. A. and Goldrick, M. (2021) 'Understanding Language Abnormalities and Associated Clinical Markers in Psychosis: The Promise of Computational Methods', *Schizophrenia Bulletin*, 47(2), pp. 344–362. doi: 10.1093/schbul/sbaa141.

Ho, C. Y. *et al.* (2010) 'Myocardial fibrosis as an early manifestation of hypertrophic cardiomyopathy', *New England Journal of Medicine*, 363(6), pp. 552–563.

Hollander, A.-C. *et al.* (2016) 'Refugee migration and risk of schizophrenia and other non-affective psychoses: cohort study of 1.3 million people in Sweden', *bmj*, 352.

Holmlund, M. *et al.* (2020) 'Customer experience management in the age of big data analytics: A strategic framework', *Journal of Business Research*, 116, pp. 356–365.

Holmlund, T. B. *et al.* (2019) 'Updating verbal fluency analysis for the 21st century: Applications for psychiatry', *Psychiatry Research*, 273, pp. 767–769.

Howes, O. D. and Kapur, S. (2009) 'The dopamine hypothesis of schizophrenia: version III--the final common pathway.', *Schizophrenia bulletin*, 35(3), pp. 549–562. doi: 10.1093/schbul/sbp006.

Howes, O., McCutcheon, R. and Stone, J. (2015) 'Glutamate and dopamine in schizophrenia: an update for the 21st century', *Journal of psychopharmacology*, 29(2), pp. 97–115.

Huber, G. and Gross, G. (1989) 'The concept of basic symptoms in schizophrenic and schizoaffective psychoses.', *Recenti progressi in medicina*, 80(12), pp. 646–652.

Ising, H. K. *et al.* (2012) 'The validity of the 16-item version of the Prodromal Questionnaire (PQ-16) to screen for ultra high risk of developing psychosis in the general

help-seeking population', *Schizophrenia bulletin*, 38(6), pp. 1288–1296.

Itil, T. *et al.* (1967) 'Effect of phencyclidine in chronic schizophrenics', *Canadian Psychiatric Association Journal*, 12(2), pp. 209–212.

Jablensky, A., McNeil, T. F. and Morgan, V. A. (2017) 'Barbara fish and a short history of the neurodevelopmental hypothesis of schizophrenia', *Schizophrenia bulletin*, 43(6), pp. 1158–1163.

Javitt, D. C. *et al.* (2001) 'Adjunctive high-dose glycine in the treatment of schizophrenia', *International Journal of Neuropsychopharmacology*, 4(4), pp. 385–391.

Javitt, D. C. (2007) 'Glutamate and schizophrenia: phencyclidine, N-methyl-d-aspartate receptors, and dopamine–glutamate interactions', *International review of neurobiology*, 78, pp. 69–108.

Javitt, D. C. *et al.* (2012) 'Has an angel shown the way? Etiological and therapeutic implications of the PCP/NMDA model of schizophrenia', *Schizophrenia bulletin*, 38(5), pp. 958–966.

Kahn, R. S. and Keefe, R. S. E. (2013) 'Schizophrenia is a cognitive illness: time for a change in focus', *JAMA psychiatry*, 70(10), pp. 1107–1112.

Kandel, E. R. *et al.* (2000) *Principles of neural science*. McGraw-hill New York.

Kang, S.-G. *et al.* (2008) 'Association study between antipsychotics-induced restless legs syndrome and polymorphisms of dopamine D1, D2, D3, and D4 receptor genes in schizophrenia', *Neuropsychobiology*, 57(1–2), pp. 49–54.

Kay, S. R., Fiszbein, A. and Opler, L. A. (1987) 'The positive and negative syndrome scale (PANSS) for schizophrenia', *Schizophrenia bulletin*, 13(2), pp. 261–276.

Keefe, R. S. E. *et al.* (2006) 'A longitudinal study of neurocognitive function in individuals at-risk for psychosis', *Schizophrenia research*, 88(1–3), pp. 26–35.

Keefe, R. S. E. *et al.* (2008) 'Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS)', *Schizophrenia research*, 102(1–3), pp. 108–115.

Kelleher, F. C. *et al.* (2012) 'Prevailing importance of the hedgehog signaling pathway and the potential for treatment advancement in sarcoma', *Pharmacology & therapeutics*, 136(2), pp. 153–168.

Kelleher, I. *et al.* (2013) 'Childhood trauma and psychosis in a prospective cohort study: cause, effect, and directionality', *American journal of psychiatry*, 170(7), pp. 734–741.

Kendler, K. S. *et al.* (1996) 'Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample: the National Comorbidity Survey', *Archives of general psychiatry*, 53(11), pp. 1022–1031.

Kew, J. N. C. and Kemp, J. A. (2005) 'Ionotropic and metabotropic glutamate receptor structure and pharmacology', *Psychopharmacology*, 179(1), pp. 4–29.

Kiss, G. *et al.* (2012) 'Quantitative analysis of pitch in speech of children with neurodevelopmental disorders', in *Thirteenth Annual Conference of the International Speech Communication Association*.

Kiss, G. and Vicsi, K. (2014) 'Physiological and cognitive status monitoring on the base of acoustic-phonetic speech parameters', in *International conference on statistical language and speech processing*. Springer, pp. 120–131.

Kita, S. (2003) *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Konopka, G. and Roberts, T. F. (2016) 'Insights into the neural and genetic basis of vocal communication', *Cell*, 164(6), pp. 1269–1276.

Kornhuber, J. *et al.* (1989) '[3H]MK-801 binding sites in postmortem brain regions of schizophrenic patients.', *Journal of neural transmission*, 77(2–3), pp. 231–236. doi: 10.1007/BF01248936.

Kraepelin, E. (1896) 'Der psychologische Versuch in der Psychiatrie', *Psychologische Arbeiten*, 1, pp. 1–91.

Krystal, J. H. *et al.* (1994) 'Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses.', *Archives of general psychiatry*, 51(3), pp. 199–214. doi: 10.1001/archpsyc.1994.03950030035004.

Kuperberg, G. R. (2007) 'Neural mechanisms of language comprehension: Challenges to syntax', *Brain research*, 1146, pp. 23–49.

Kuperberg, G. R. *et al.* (2008) 'Functional magnetic resonance imaging reveals neuroanatomical dissociations during semantic integration in schizophrenia', *Biological psychiatry*, 64(5), pp. 407–418.

Kuperberg, G. R. (2010a) 'Language in schizophrenia part 1: an introduction', *Language and linguistics compass*, 4(8), pp. 576–589.

Kuperberg, G. R. (2010b) 'Language in schizophrenia Part 2: What can psycholinguistics

bring to the study of schizophrenia… and vice versa?', *Language and linguistics compass*, 4(8), pp. 590–604.

Lakens, D. *et al.* (2018) 'Justify your alpha', *Nature Human Behaviour*, 2(3), pp. 168–171.

Landauer, T. K., Foltz, P. W. and Laham, D. (1998) 'An introduction to latent semantic analysis', *Discourse processes*, 25(2–3), pp. 259–284.

Lau, S. K. P. *et al.* (2013) 'Delayed induction of proinflammatory cytokines and suppression of innate antiviral response by the novel Middle East respiratory syndrome coronavirus: implications for pathogenesis and treatment', *Journal of General Virology*, 94(12), pp. 2679–2690.

Laursen, T. M., Nordentoft, M. and Mortensen, P. B. (2014) 'Excess early mortality in schizophrenia', *Annu Rev Clin Psychol*, 10(1), pp. 425–448.

Lawrence, R. E., First, M. B. and Lieberman, J. A. (2015) 'Schizophrenia and other psychoses', *Psychiatry*, pp. 791–856.

Ledonne, A. and Mercuri, N. B. (2017) 'Current concepts on the physiopathological relevance of dopaminergic receptors', *Frontiers in cellular neuroscience*, 11, p. 27.

Legge, S. E. *et al.* (2019) 'Association of genetic liability to psychotic experiences with neuropsychotic disorders and traits', *JAMA psychiatry*, 76(12), pp. 1256–1265.

Lennes, M. *et al.* (2016) 'Comparing pitch distributions using Praat and R', *Phonetician*.

Lewis, D. A. *et al.* (2012) 'Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia', *Trends in neurosciences*, 35(1), pp. 57–67.

Li, P., L Snyder, G. and E Vanover, K. (2016) 'Dopamine targeting drugs for the treatment of schizophrenia: past, present and future', *Current topics in medicinal chemistry*, 16(29), pp. 3385–3403.

Li, T. *et al.* (2013) 'FoxP2 is significantly associated with schizophrenia and major depression in the Chinese Han population', *The World Journal of Biological Psychiatry*, 14(2), pp. 146–150.

Li, X. *et al.* (2019) 'Altered topological characteristics of morphological brain network relate to language impairment in high genetic risk subjects and schizophrenia patients', *Schizophrenia Research*, 208, pp. 338–343.

Lim, A. *et al.* (2016) 'Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders', *Schizophrenia research*, 176(2–3), pp.

493–499.

Lin, A. *et al.* (2011) ‘Neurocognitive predictors of functional outcome two to 13 years after identification as ultra-high risk for psychosis’, *Schizophrenia research*, 132(1), pp. 1–7.

Linn, D. and Hemmer, L. (2011) ‘English Language Learner Disproportionality in Special Education: Implications for the Scholar-Practitioner.’, *Journal of Educational Research and Practice*, 1(1), pp. 70–80.

Lisman, J. (2012) ‘Excitation, inhibition, local oscillations, or large-scale loops: what causes the symptoms of schizophrenia?’, *Current opinion in neurobiology*, 22(3), pp. 537–544.

Lo, L. E. *et al.* (2020) ‘Risk of schizophrenia in relatives of individuals affected by schizophrenia: a meta-analysis’, *Psychiatry Research*, 286, p. 112852.

Loebel, A. D. *et al.* (1992) ‘Duration of psychosis and outcome in first-episode schizophrenia.’, *The American journal of psychiatry*.

Loewy, R. L. *et al.* (2005) ‘The prodromal questionnaire (PQ): preliminary validation of a self-report screening measure for prodromal and psychotic syndromes’, *Schizophrenia research*, 79(1), pp. 117–125.

Low, D. M., Bentley, K. H. and Ghosh, S. S. (2020) ‘Automated assessment of psychiatric disorders using speech: A systematic review’, *Laryngoscope Investigative Otolaryngology*, 5(1), pp. 96–116. doi: 10.1002/lio2.354.

Luby, E. D. *et al.* (1962) ‘Model psychoses and schizophrenia’, *American Journal of Psychiatry*, 119(1), pp. 61–67.

Luck, S. J. and Gold, J. M. (2008) ‘The construct of attention in schizophrenia’, *Biological psychiatry*, 64(1), pp. 34–39.

Lysaker, P. and Bell, M. (1995) ‘Negative symptoms and vocational impairment in schizophrenia: Repeated measurements of work performance over six months’, *Acta Psychiatrica Scandinavica*, 91(3), pp. 205–208.

Mackenzie, C. S., Gekoski, W. L. and Knox, V. J. (2006) ‘Age, gender, and the underutilization of mental health services: The influence of help-seeking attitudes’, *Aging and mental health*, 10(6), pp. 574–582.

Markov, V. *et al.* (2009) ‘Genetic variation in schizophrenia-risk-gene dysbindin 1

modulates brain activation in anterior cingulate cortex and right temporal gyrus during language production in healthy individuals', *Neuroimage*, 47(4), pp. 2016–2022.

Marsman, A. *et al.* (2013) 'Glutamate in schizophrenia: a focused review and meta-analysis of 1H-MRS studies', *Schizophrenia bulletin*, 39(1), pp. 120–129.

Martel, J. C. and Gatti McArthur, S. (2020) 'Dopamine receptor subtypes, physiology and pharmacology: new ligands and concepts in schizophrenia', *Frontiers in pharmacology*, 11, p. 1003.

Martínez-Sánchez, F. *et al.* (2015) 'Can the acoustic analysis of expressive prosody discriminate schizophrenia?', *The Spanish journal of psychology*, 18.

Martínez-Sánchez, F. *et al.* (2017) 'Speech rhythm alterations in Spanish-speaking individuals with Alzheimer's disease', *Aging, Neuropsychology, and Cognition*, 24(4), pp. 418–434.

Matsumoto, K. *et al.* (2013) 'Frequency and neural correlates of pauses in patients with formal thought disorder', *Frontiers in Psychiatry*, 4, p. 127.

McDonald, M. *et al.* (2019) 'Using online screening in the general population to detect participants at clinical high-risk for psychosis', *Schizophrenia bulletin*, 45(3), pp. 600–609.

McGlashan, T. H. *et al.* (2001) 'Instrument for the assessment of prodromal symptoms and states', in *Early intervention in psychotic disorders*. Springer, pp. 135–149.

McGlashan, T., Walsh, B. and Woods, S. (2010) *The psychosis-risk syndrome: handbook for diagnosis and follow-up*. Oxford University Press.

McGorry, P. D. *et al.* (1995) 'The prevalence of prodromal features of schizophrenia in adolescence: a preliminary survey', *Acta Psychiatrica Scandinavica*, 92(4), pp. 241–249.

McGorry, P. D. *et al.* (2018) 'Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry', *World Psychiatry*, 17(2), pp. 133–142.

McGrath, J. *et al.* (2008) 'Schizophrenia: a concise overview of incidence, prevalence, and mortality', *Epidemiologic reviews*, 30(1), pp. 67–76.

McGrath, J. J. *et al.* (2003) 'The neurodevelopmental hypothesis of schizophrenia: a review of recent developments', *Annals of medicine*, 35(2), pp. 86–93.

McGuire, P. *et al.* (2015) 'Can neuroimaging be used to predict the onset of psychosis?', *The Lancet Psychiatry*, 2(12), pp. 1117–1122.

McGurk, S. R. *et al.* (2004) ‘Cognitive functioning predicts outpatient service utilization in schizophrenia’, *Mental health services research*, 6(3), pp. 185–188.

McGurk, S. R. *et al.* (2007) ‘A meta-analysis of cognitive remediation in schizophrenia’, *American Journal of Psychiatry*, 164(12), pp. 1791–1802.

McGurk, S. R. and Mueser, K. T. (2004) ‘Cognitive functioning, symptoms, and work in supported employment: a review and heuristic model’, *Schizophrenia research*, 70(2–3), pp. 147–173.

Melinder, M. R. D. and Barch, D. M. (2003) ‘The influence of a working memory load manipulation on language production in schizophrenia’, *Schizophrenia Bulletin*, 29(3), pp. 473–485.

Ménard, R., Skachko, S. and Pannekoucke, O. (2021) ‘Numerical discretization causing error variance loss and the need for inflation’, *Quarterly Journal of the Royal Meteorological Society*, 147(740), pp. 3498–3520.

Mendoza, E. and Carballo, G. (1998) ‘Acoustic analysis of induced vocal stressby means of cognitive workload tasks’, *Journal of Voice*, 12(3), pp. 263–273.

Mesholam-Gately, R. I. *et al.* (2009) ‘Neurocognition in first-episode schizophrenia: a meta-analytic review.’, *Neuropsychology*, 23(3), p. 315.

Michel, C. *et al.* (2018) ‘Course of clinical high-risk states for psychosis beyond conversion’, *European archives of psychiatry and clinical neuroscience*, 268(1), pp. 39–48.

Millan, M. J. *et al.* (2016) ‘Altering the course of schizophrenia: progress and perspectives’, *Nature Reviews Drug Discovery*, 15(7), pp. 485–515.

Miller, B. J. *et al.* (2011) ‘Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects’, *Biological psychiatry*, 70(7), pp. 663–671.

Miller, P. *et al.* (2002) ‘Schizotypal components in people at high risk of developing schizophrenia: early findings from the Edinburgh High-Risk Study’, *The British Journal of Psychiatry*, 180(2), pp. 179–184.

Miller, T. J. *et al.* (1999) ‘Symptom assessment in schizophrenic prodromal states’, *Psychiatric quarterly*, 70(4), pp. 273–287.

Millman, Z. B. *et al.* (2019) ‘The critical need for help-seeking controls in clinical high-risk research’, *Clinical Psychological Science*, 7(6), pp. 1171–1189.

Misiak, B. *et al.* (2016) 'Childhood traumatic events and types of auditory verbal hallucinations in first-episode schizophrenia patients', *Comprehensive psychiatry*, 66, pp. 17–22.

Misiak, B. *et al.* (2018) 'Cytokines and C-reactive protein alterations with respect to cognitive impairment in schizophrenia and bipolar disorder: a systematic review', *Schizophrenia research*, 192, pp. 16–29.

Moghaddam, B. (2003) 'Bringing order to the glutamate chaos in schizophrenia', *Neuron*, 40(5), pp. 881–884.

Moore, T. H. M. *et al.* (2007) 'Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review', *The Lancet*, 370(9584), pp. 319–328.

Moore, T. M. *et al.* (2015) 'Psychometric properties of the Penn Computerized Neurocognitive Battery.', *Neuropsychology*, 29(2), p. 235.

Morgan, C. J. *et al.* (2017) 'Thought disorder in schizophrenia and bipolar disorder probands, their relatives, and nonpsychiatric controls', *Schizophrenia bulletin*, 43(3), pp. 523–535.

Morgan, S. E. *et al.* (2021) 'Natural Language Processing markers in first episode psychosis and people at clinical high-risk', *Translational psychiatry*, 11(1), pp. 1–9.

Morrison, A. P. *et al.* (2012) 'Early detection and intervention evaluation for people at risk of psychosis: multisite randomised controlled trial', *Bmj*, 344.

Mota, C. and Santos, D. (2015) 'Emotions in natural language: a broad-coverage perspective'.

Mota, N. B., Copelli, M. and Ribeiro, S. (2017) 'Thought disorder measured as random speech structure classifies negative symptoms and schizophrenia diagnosis 6 months in advance', *npj Schizophrenia*, 3(1), pp. 1–10.

Mota, P. *et al.* (2012) 'Natural language understanding: From laboratory predictions to real interactions', in *International Conference on Text, Speech and Dialogue*. Springer, pp. 640–647.

Nakazawa, D. *et al.* (2012) 'Abundant neutrophil extracellular traps in thrombus of patient with microscopic polyangiitis', *Frontiers in immunology*, 3, p. 333.

Nakazawa, K., Jeevakumar, V. and Nakao, K. (2017) 'Spatial and temporal boundaries of NMDA receptor hypofunction leading to schizophrenia', *NPJ schizophrenia*, 3(1), pp. 1–

11.

Nelson, B. *et al.* (2013) ‘Long-term follow-up of a group at ultra high risk (“prodromal”) for psychosis: the PACE 400 study’, *JAMA psychiatry*, 70(8), pp. 793–802.

Nielsen, P. R., Meyer, U. and Mortensen, P. B. (2016) ‘Individual and combined effects of maternal anemia and prenatal infection on risk for schizophrenia in offspring’, *Schizophrenia Research*, 172(1–3), pp. 35–40.

Nielsen, S. M. *et al.* (2017) ‘Association between alcohol, cannabis, and other illicit substance abuse and risk of developing schizophrenia: a nationwide population based register study’, *Psychological medicine*, 47(9), pp. 1668–1677.

Ødegaard, O. (1932) ‘Migration and insanity. A study of mental disease among the Norwegian population of Minnesota’, *Acta Psych Neurol Scand*, 4, pp. 201–232.

Olney, J. W. and Farber, N. B. (1995) ‘Glutamate receptor dysfunction and schizophrenia’, *Archives of general psychiatry*, 52(12), pp. 998–1007.

Olney, J. W., Newcomer, J. W. and Farber, N. B. (1999) ‘NMDA receptor hypofunction model of schizophrenia.’, *Journal of psychiatric research*, 33(6), pp. 523–533. doi: 10.1016/s0022-3956(99)00029-1.

Olsen, K. A. and Rosenbaum, B. (2006) ‘Prospective investigations of the prodromal state of schizophrenia: assessment instruments’, *Acta Psychiatrica Scandinavica*, 113(4), pp. 273–282.

Organization, W. H. (2013) *Building back better: sustainable mental health care after emergencies*. World Health Organization.

Organization, W. H. (2018) *WHO recommendations on intrapartum care for a positive childbirth experience*. World Health Organization.

Orygen (2015) ‘The CAARMS: Assessing Young People at Ultra High Risk of Psychosis’.

Van Os, J. (2009) ‘‘Salience syndrome’replaces ‘schizophrenia’in DSM-V and ICD-11: psychiatry’s evidence-based entry into the 21st century?’, *Acta Psychiatrica Scandinavica*, 120(5), pp. 363–372.

Palaniyappan, L. (2021) ‘More than a biomarker: could language be a biosocial marker of psychosis?’, *npj Schizophrenia*, 7(1), pp. 1–5.

Pantelis, C. *et al.* (2003) ‘Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison’, *The Lancet*, 361(9354),

pp. 281–288.

Parola, A. *et al.* (2020) ‘Voice patterns in schizophrenia: A systematic review and Bayesian meta-analysis’, *Schizophrenia Research*, 216, pp. 24–40. doi: 10.1016/j.schres.2019.11.031.

Pennebaker, J. W., Mehl, M. R. and Niederhoffer, K. G. (2003) ‘Psychological aspects of natural language use: Our words, our selves’, *Annual review of psychology*, 54(1), pp. 547–577.

Peralta, V. and Cuesta, M. J. (1999) ‘Dimensional structure of psychotic symptoms: an item-level analysis of SAPS and SANS symptoms in psychotic disorders’, *Schizophrenia research*, 38(1), pp. 13–26.

Pilowsky, L. S. *et al.* (2006) ‘First in vivo evidence of an NMDA receptor deficit in medication-free schizophrenic patients’, *Molecular psychiatry*, 11(2), pp. 118–119.

Pinheiro, A. P. *et al.* (2016) ‘Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia’, *Cognitive neuropsychiatry*, 21(4), pp. 335–353.

Pinheiro, A. P. *et al.* (2017) ‘Emotional self–other voice processing in schizophrenia and its relationship with hallucinations: ERP evidence’, *Psychophysiology*, 54(9), pp. 1252–1265.

Pinker, S. (2003) *The language instinct: How the mind creates language*. Penguin UK.

Pletcher, M. J. and Pignone, M. (2011) ‘Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact’, *Circulation*, 123(10), pp. 1116–1124.

Pocklington, A. J. *et al.* (2015) ‘Novel Findings from CNVs Implicate Inhibitory and Excitatory Signaling Complexes in Schizophrenia.’, *Neuron*, 86(5), pp. 1203–1214. doi: 10.1016/j.neuron.2015.04.022.

Poulton, R. *et al.* (2000) ‘Children’s self-reported psychotic symptoms and adult schizophreniform disorder: a 15-year longitudinal study’, *Archives of general psychiatry*, 57(11), pp. 1053–1058.

Price, C. J. (2010) ‘The anatomy of language: a review of 100 fMRI studies published in 2009’, *Annals of the new York Academy of Sciences*, 1191(1), pp. 62–88.

Purcell, S. *et al.* (2009) ‘Common polygenic variation contributes to risk of schizophrenia and bipolar disorder’, , *Nature*, 460(7256), pp. 748–752.

Quené, H., Persoon, I. and de Jong, N. (2011) 'Syllable Nuclei v2 [Praat Script]. Version 28 Feb 2011'.

Rapcan, V. *et al.* (2010) 'Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia', *Medical Engineering and Physics*, 32(9), pp. 1074–1079. doi: 10.1016/j.medengphy.2010.07.013.

Rezaii, N., Walker, E. and Wolff, P. (2019) 'A machine learning approach to predicting psychosis using semantic density and latent content analysis', *npj Schizophrenia*, 5(1). doi: 10.1038/s41537-019-0077-9.

Riecher-Rössler, A. *et al.* (2008) 'The Basel Screening Instrument for Psychosis (BSIP): development, structure, reliability and validity', *Fortschritte der Neurologie-Psychiatrie*, 76(4), pp. 207–216.

Ripke, S. *et al.* (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511(7510), pp. 421–427. doi: 10.1038/nature13595.

Robbins, T. W. and Arnsten, A. (2009) 'The neuropsychopharmacology of fronto-executive function: monoaminergic modulation', *Annual review of neuroscience*, 32, p. 267.

Roche, E. *et al.* (2015) 'The epidemiology and associated phenomenology of formal thought disorder: a systematic review', *Schizophrenia bulletin*, 41(4), pp. 951–962.

Rosen, S. (1992) 'Temporal information in speech: acoustic, auditory and linguistic aspects', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 336(1278), pp. 367–373.

Rosso, I. M. *et al.* (2000) 'Obstetric risk factors for early-onset schizophrenia in a Finnish birth cohort', *American Journal of Psychiatry*, 157(5), pp. 801–807.

Rothman, D. H., Hayes, J. M. and Summons, R. E. (2003) 'Dynamics of the Neoproterozoic carbon cycle', *Proceedings of the National Academy of Sciences*, 100(14), pp. 8124–8129.

Ruderfer, D. M. *et al.* (2016) 'Polygenic overlap between schizophrenia risk and antipsychotic response: a genomic medicine approach', *The Lancet Psychiatry*, 3(4), pp. 350–357.

Saha, S. *et al.* (2005) 'A systematic review of the prevalence of schizophrenia', *PLoS medicine*, 2(5), p. e141.

Sans-Sansa, B. *et al.* (2013) 'Association of formal thought disorder in schizophrenia with structural brain abnormalities in language-related cortical regions', *Schizophrenia research*, 146(1–3), pp. 308–313.

Schlosser, D. A. *et al.* (2012) 'Recovery from an at-risk state: clinical and functional outcomes of putatively prodromal youth who do not develop psychosis', *Schizophrenia bulletin*, 38(6), pp. 1225–1233.

Schultze-Lutter, F. *et al.* (2007) 'Basic symptoms in early psychotic and depressive disorders', *The British Journal of Psychiatry*, 191(S51), pp. s31–s37.

Schultze-Lutter, F. *et al.* (2015) 'EPA guidance on the early detection of clinical high risk states of psychoses', *European Psychiatry*, 30(3), pp. 405–416.

Schultze-Lutter, F. *et al.* (2016) 'Revisiting the basic symptom concept: toward translating risk symptoms for psychosis into neurobiological targets', *Frontiers in psychiatry*, 7, p. 9.

Schwartz, T. L., Sachdeva, S. and Stahl, S. M. (2012) 'Glutamate neurocircuitry: theoretical underpinnings in schizophrenia', *Frontiers in pharmacology*, 3, p. 195.

Seidenfaden, D. *et al.* (2017) 'The relationship between self-reported childhood adversities, adulthood psychopathology and psychological stress markers in patients with schizophrenia', *Comprehensive psychiatry*, 72, pp. 48–55.

Sheehan, D. V *et al.* (1998) 'The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10', *Journal of clinical psychiatry*, 59(20), pp. 22–33.

Shih, R. A. *et al.* (2011) 'Childhood neuromotor dysfunction in schizophrenia patients and their unaffected siblings: a prospective cohort study', *Schizophrenia bulletin*, 70(2), pp. 367–378. doi: 10.1080/09540260400014401.

Shmueli, G. (2010) 'To explain or to predict?', *Statistical science*, 25(3), pp. 289–310.

Sichlinger, L. *et al.* (2019) 'Clinical correlates of aberrant conversational turn-taking in youth at clinical high-risk for psychosis', *Schizophrenia research*, 204, p. 419.

Sideli, L. *et al.* (2012) 'Do child abuse and maltreatment increase risk of schizophrenia?', *Psychiatry investigation*, 9(2), p. 87.

Sigman, M. and Cecchi, G. A. (2002) 'Global organization of the Wordnet lexicon', *Proceedings of the National Academy of Sciences*, 99(3), pp. 1742–1747.

Simon, A. E. *et al.* (2011) 'Ultra high-risk state for psychosis and non-transition: a

systematic review', *Schizophrenia research*, 132(1), pp. 8–17.

Simon, A. E. and Umbricht, D. (2010) 'High remission rates from an initial ultra-high risk state for psychosis', *Schizophrenia research*, 116(2–3), pp. 168–172.

Sisti, D. A. and Calkins, M. E. (2016) 'Psychosis risk: what is it and how should we talk about it?', *AMA Journal of Ethics*, 18(6), pp. 624–632.

Sobin, C. and Alpert, M. (1999) 'Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy', *Journal of psycholinguistic research*, 28(4), pp. 347–365.

Sørensen, H. J. *et al.* (2014) 'Population impact of familial and environmental risk factors for schizophrenia: a nationwide study', *Schizophrenia research*, 153(1–3), pp. 214–219.

Sørensen, J. D. (2009) 'Framework for risk-based planning of operation and maintenance for offshore wind turbines', *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, 12(5), pp. 493–506.

Spencer, T. J. *et al.* (2021) 'Lower speech connectedness linked to incidence of psychosis in people at clinical high risk', *Schizophrenia research*, 228, pp. 493–501.

Srinivasan, N. (2018) 'Acoustic analysis of English vowels by young Spanish-English bilingual language learners'. The George Washington University.

Stanislawski, E. R. *et al.* (2021) 'Negative symptoms and speech pauses in youths at clinical high risk for psychosis', *npj Schizophrenia*, 7(1), pp. 1–3.

Startup, M., Jackson, M. C. and Bendix, S. (2002) 'The concurrent validity of the Global Assessment of Functioning (GAF)', *British Journal of Clinical Psychology*, 41(4), pp. 417–422.

Stassen, H. H., Kuny, S. and Hell, D. (1998) 'The speech analysis approach to determining onset of improvement under antidepressants', *European Neuropsychopharmacology*, 8(4), pp. 303–310.

Stone, J. M. *et al.* (2008) 'Relationship between ketamine-induced psychotic symptoms and NMDA receptor occupancy—a [123I] CNS-1261 SPET study', *Psychopharmacology*, 197(3), pp. 401–408.

Susser, E., Hoek, H. W. and Brown, A. (1998) 'Neurodevelopmental disorders after prenatal famine: the story of the Dutch Famine Study', *American journal of epidemiology*, 147(3), pp. 213–216.

Tabachnick, B. G. and Fidell, L. S. (1996) *SPSS for Windows workbook to accompany*

large sample examples of using multivariate statistics. HarperCollins College Publishers.

Tahir, Y. *et al.* (2019) 'Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia', *PLoS One*, 14(4), p. e0214314.

Tan, E. J. *et al.* (2021) 'Investigating the diagnostic utility of speech patterns in schizophrenia and their symptom associations', *Schizophrenia Research*, 238, pp. 91–98.

Tan, E. J. and Rossell, S. L. (2019) 'Language comprehension and neurocognition independently and concurrently contribute to formal thought disorder severity in schizophrenia', *Schizophrenia research*, 204, pp. 133–137.

Tanaka, M. *et al.* (2008) 'Hyperglycosylation and reduced GABA currents of mutated GABRB3 polypeptide in remitting childhood absence epilepsy', *The American Journal of Human Genetics*, 82(6), pp. 1249–1261.

Tavakoli, P. and Skehan, P. (2005) 'Strategic planning, task structure, and performance testing', *Planning and task performance in a second language*, 239273.

Team, A. (2021) 'Audacity®: free audio editor and recorder [Computer application]. Version 3.0. 0'.

Teixeira, J. P., Oliveira, C. and Lopes, C. (2013) 'Vocal acoustic analysis–jitter, shimmer and hnr parameters', *Procedia Technology*, 9, pp. 1112–1122.

Thapar, A. and Cooper, M. (2013) 'Copy number variation: what is it and what has it told us about child psychiatric disorders?', *Journal of the American Academy of Child and Adolescent Psychiatry*, 52(8), p. 772.

Thermenos, H. W. *et al.* (2013) 'Altered language network activity in young people at familial high-risk for schizophrenia', *Schizophrenia research*, 151(1–3), pp. 229–237.

Thornberg, S. A. and Saklad, S. R. (1996) 'A review of NMDA receptors and the phencyclidine model of schizophrenia', *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, 16(1), pp. 82–93.

Titze, I. R. (1989) 'Physiologic and acoustic differences between male and female voices', *The Journal of the Acoustical Society of America*, 85(4), pp. 1699–1707.

Tolosa, A. *et al.* (2010) 'FOXP2 gene and language impairment in schizophrenia: association and epigenetic studies', *BMC medical genetics*, 11(1), pp. 1–8.

Trémeau, F. *et al.* (2013) 'Inpatients with schizophrenia report impaired situational motivation but intact global and social motivation', *Psychiatry research*, 210(1), pp. 43–

49.

Uhlhaas, P. J. *et al.* (2010) 'Neural synchrony and the development of cortical networks', *Trends in cognitive sciences*, 14(2), pp. 72–80.

Uhlhaas, P. J. (2013) 'Dysconnectivity, large-scale networks and neuronal dynamics in schizophrenia', *Current opinion in neurobiology*, 23(2), pp. 283–290.

Uhlhaas, P. J. *et al.* (2017) 'The youth mental health risk and resilience study (YouR-Study)', *BMC psychiatry*, 17(1), pp. 1–8.

Uhlhaas, P. J. and Singer, W. (2010) 'Abnormal neural oscillations and synchrony in schizophrenia', *Nature reviews neuroscience*, 11(2), pp. 100–113.

Varese, F. *et al.* (2012) 'Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective-and cross-sectional cohort studies', *Schizophrenia bulletin*, 38(4), pp. 661–671.

Walter, A. *et al.* (2016) 'Hippocampal volume in subjects at clinical high-risk for psychosis: A systematic review and meta-analysis', *Neuroscience & Biobehavioral Reviews*, 71, pp. 680–690.

Walther, S. *et al.* (2015) 'Nonverbal social communication and gesture control in schizophrenia', *Schizophrenia Bulletin*, 41(2), pp. 338–345.

Wassef, A., Baker, J. and Kochan, L. D. (2003) 'GABA and schizophrenia: a review of basic science and clinical studies', *Journal of clinical psychopharmacology*, 23(6), pp. 601–640.

Weickert, C. S. *et al.* (2013) 'Molecular evidence of N-methyl-D-aspartate receptor hypofunction in schizophrenia.', *Molecular psychiatry*, 18(11), pp. 1185–1192. doi: 10.1038/mp.2012.137.

Wennerstrom, A. (2001) *The music of everyday speech: Prosody and discourse analysis*. Oxford University Press.

De Wit, S. *et al.* (2014) 'Adolescents at ultra-high risk for psychosis: long-term outcome of individuals who recover from their at-risk state', *European Neuropsychopharmacology*, 24(6), pp. 865–873.

Wykes, T. *et al.* (2011) 'A meta-analysis of cognitive remediation for schizophrenia: methodology and effect sizes', *American Journal of Psychiatry*, 168(5), pp. 472–485.

Xu, B. *et al.* (2008) 'Strong association of de novo copy number mutations with sporadic

schizophrenia', *Nature genetics*, 40(7), pp. 880–885.

Yalincetin, B. *et al.* (2017) 'Formal thought disorder in schizophrenia and bipolar disorder: A systematic review and meta-analysis', *Schizophrenia research*, 185, pp. 2–8.

Yildirim, S. *et al.* (2004) 'An acoustic study of emotions expressed in speech', in *Eighth International Conference on Spoken Language Processing*.

Yung, A. *et al.* (2002) 'Comprehensive assessment of at-risk mental states (CAARMS)', *Melbourne, Australia, University of Melbourne, Department of Psychiatry, Personal Assessment and Crisis Evaluation Clinic*.

Yung, A. R. *et al.* (2005) 'Mapping the onset of psychosis: the comprehensive assessment of at-risk mental states', *Australian & New Zealand Journal of Psychiatry*, 39(11–12), pp. 964–971.

Yung, A. R. *et al.* (2019) 'Persistent negative symptoms in individuals at Ultra High Risk for psychosis', *Schizophrenia research*, 206, pp. 355–361.

Yung, A. R. and McGorry, P. D. (1996) 'The prodromal phase of first-episode psychosis: past and current conceptualizations', *Schizophrenia bulletin*, 22(2), pp. 353–370.

Yung, A. R. and McGorry, P. D. (2007) 'Prediction of psychosis: setting the stage', *The British Journal of Psychiatry*, 191(S51), pp. s1–s8.

Zhang, Y. and Jiang, J. J. (2008) 'Acoustic analyses of sustained and running voices from patients with laryngeal pathologies', *Journal of Voice*, 22(1), pp. 1–9.

Zhang, Z. (2016) 'Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model', *The Journal of the Acoustical Society of America*, 139(4), pp. 1493–1507.

Zhou, Y. and Danbolt, N. C. (2014) 'Glutamate as a neurotransmitter in the healthy brain', *Journal of neural transmission*, 121(8), pp. 799–817.

Zwetsch, I. C. *et al.* (2006) 'Digital signal processing in the differential diagnosis of benign larynx diseases [Abstract in English]', *Scientia Medica*, 16(3), pp. 109–114.