Deep Clinical Phenotyping of Schizophrenia Spectrum Disorders Using Data-Driven Methods

Habtewold, Tesfa Dejenie; Hao, Jiasi; Liemburg, Edith J.; Baştürk, Nalan; Bruggeman, Richard; Alizadeh, Behrooz Z.

Link to publication in University of Groningen/UMCG research database

Brief Report

Deep Clinical Phenotyping of Schizophrenia Spectrum Disorders Using Data-Driven Methods: Marching towards Precision Psychiatry

Tesfa Dejenie Habtewold [1,*] , **Jiasi Hao** [1] , **Edith J. Liemburg** [2] , **Nalan Baştürk** [3] , **Richard Bruggeman** [2,4] and **Behrooz Z. Alizadeh** [1,*]

1 Department of Epidemiology, University Medical Center Groningen, University of Groningen, 9700 RB Groningen, The Netherlands; j.hao@umcg.nl

2 Department of Psychiatry, University Medical Center Groningen, University Center for Psychiatry, Rob Giel Research Center, University of Groningen, 9700 RB Groningen, The Netherlands; e.j.liemburg@umcg.nl (E.J.L.); r.bruggeman@umcg.nl (R.B.)

3 Department of Quantitative Economics, School of Business and Economics, Maastricht University, 6200 MD Maastricht, The Netherlands; n.basturk@maastrichtuniversity.nl

4 Department of Clinical and Developmental Neuropsychology, Faculty of Behavioural and Social Sciences, University of Groningen, 9700 RB Groningen, The Netherlands

* Correspondence: t.d.habtewold@umcg.nl (T.D.H.); b.z.alizadeh@umcg.nl (B.Z.A.); Tel.: +31-50-361-0738 (T.D.H. & B.Z.A.)

Abstract: Heterogeneity is the main challenge in the traditional classification of mental disorders, including schizophrenia spectrum disorders (SSD). This can be partly attributed to the absence of objective diagnostic criteria and the multidimensional nature of symptoms and their associated factors. This article provides an overview of findings from the Genetic Risk and Outcome of Psychosis (GROUP) cohort study on the deep clinical phenotyping of schizophrenia spectrum disorders targeting positive and negative symptoms, cognitive impairments and psychosocial functioning. Three to four latent subtypes of positive and negative symptoms were identified in patients, siblings and controls, whereas four to six latent cognitive subtypes were identified. Five latent subtypes of psychosocial function—multidimensional social inclusion and premorbid adjustment—were also identified in patients. We discovered that the identified subtypes had mixed profiles and exhibited stable, deteriorating, relapsing and ameliorating longitudinal courses over time. Baseline positive and negative symptoms, premorbid adjustment, psychotic-like experiences, health-related quality of life and PRS_{SCZ} were found to be the strong predictors of the identified subtypes. Our findings are comprehensive, novel and of clinical interest for precisely identifying high-risk population groups, patients with good or poor disease prognosis and the selection of optimal intervention, ultimately fostering precision psychiatry by tackling diagnostic and treatment selection challenges pertaining to heterogeneity.

Keywords: schizophrenia spectrum disorder; psychosis; phenotype; deep phenotyping; cognitive impairment; data-driven methods; precision psychiatry

1. Introduction

Schizophrenia is one of the most common and severe psychiatric disorders, with a lifetime prevalence of approximately 1% [1]. Globally, schizophrenia cases almost doubled from 13.1 million in 1990 to 20.9 million in 2016, and it has contributed to 13.4 million years of life lived with disability, which brings a substantial burden on patients, families, communities and healthcare systems [2].

The first episode of schizophrenia usually occurs in late adolescence or early adulthood [1]. The hallmark manifestations of schizophrenia include positive (i.e., delusions,

hallucinations and disorganized behaviour) and negative (i.e., blunted affect and asociality) symptoms [3]. Positive symptoms are the main presenting complaints and reason that patients seek treatment [4]. They are characterized by relapses and remission, which are often managed through antipsychotic treatment [4]. Negative symptoms are more stable, persistent and difficult to treat using pharmacological and non-pharmacological interventions [5]. Negative symptoms are believed to often occur prior to the onset of schizophrenia [5].

Cognitive deficits are also common symptoms observed in up to 90% of patients with schizophrenia, although they are not yet included as a diagnostic criterion of schizophrenia in DSM [6,7]. The most affected cognitive functions are episodic and working memory, attention, verbal fluency, executive function, problem-solving, processing speed and social cognition [8]. Cognitive deficits are stable and considered heritable (h^2 = 50%) traits of schizophrenia [9]. Similar to negative symptoms, cognitive symptoms could also occur before the onset of schizophrenia [5] and barely respond to antipsychotic treatment [10].

Positive, negative and cognitive symptoms have pervasive impacts on the psychosocial functioning, occupational and academic performance, and quality of life of a patient [5,7]. Furthermore, it is noteworthy to indicate that these phenotypes are also diversely present in siblings due to shared genetic and non-genetic factors with their probands [6].

The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) is the most widespread nosographic manual used by international scientific and clinical communities to diagnose mental disorders [3,11]. Clinicians subjectively interview patients and diagnose based on the set of criteria listed in the DSM [11]. However, the DSM has been criticized for the inability of diagnostic categories to map cleanly onto either biology or outcome, due to its overreliance on positive symptoms and negligence of affective, negative and cognitive symptoms [12]. For example, this leads to up to a 50% misdiagnosis rate in schizophrenia [13]. The high rate of misdiagnosis and the consequent misclassifications add a considerable burden due to ineffective interventions, poor disease progression, resource wastage and inconsistency in research reports [13]. Moreover, attempts to identify objective diagnostic biomarkers, such as genetic, proteomic and immunological markers, for schizophrenia have not yet been successful because of methodological shortcomings, professional indecision and conceptual contradictions [14]. Lastly, both psychiatric and non-psychiatric comorbidities complicate the treatment and prediction of schizophrenia outcomes [15]. Therefore, the diagnostic methods require reconsideration for a better translation of evidence into accurate and consistent clinical practice.

Several solutions have been proposed for tackling the aforementioned issues. First, DSM-5 has attempted to shift away from categorical diagnostics and adopted a dimensional approach that emphasizes the underlying specific symptoms or symptom clusters and their relationship with each other [3]. DSM-5 introduced the concept of schizophrenia spectrum disorders (SSD), encompassing various diagnoses, including schizophrenia, although in practice, diagnoses remain exclusive rather than to be seen as spectrum of disorders [3]. Another thriving solution is the use of variable and patient-centred data-driven approaches that classify psychiatric disorders to find more homogeneous subgroups of individuals in the exploitation of multidimensional symptom profiles [16].

Up till now, the narratives have shown the inherent phenotypic heterogeneity of SSD, which consequently and urgently calls for deep clinical phenotyping using large-scale studies with the possibility of in-depth exploration of disease subtypes and characterization. Previous epidemiological studies on SSD were hampered by limited sample size, limited use of omics data, and short duration of follow-up [2]. Of interest, no study comprehensively examined phenotypes of SSD and underlying factors in patients, unaffected siblings and controls.

Detecting phenotypic subgroups of patients suffering from complex diseases is an important element of precision medicine and could provide support for the early detection of deteriorating patients, prediction of individualised and customised treatment, and prevention strategies for different phenotypic groups, which ultimately results in enhanced

treatment outcome [17,18]. The in-depth analysis of phenotype plays a key role in clinical practice and medical research [18]. This article provides an overview of findings from eight Genetic Risk and Outcome of Psychosis (GROUP) cohort studies on the deep clinical phenotyping of schizophrenia spectrum disorders targeting positive and negative symptoms [19,20], cognitive impairments [21–23] and psychosocial functioning [24–26]. Briefly, we present identified latent subtypes (i.e., subphenotypes), their longitudinal course or profiles of subtypes and important predictors. Finally, we present future research directions to extend this line of research.

2. Methods and Materials

2.1. Study Population

Patients, siblings and controls participated in the Genetic Risk and Outcome of Psychosis (GROUP) cohort study. GROUP is a six-year, longitudinal cohort study in the Netherlands [27]. Patients were included if they were diagnosed with a schizophrenia spectrum disorder (53.8% schizophrenia), were aged between 16 and 50 years old, had a good command of the Dutch language and were willing and capable of giving written informed consent. Siblings and controls were included if they met the above criteria and had no known lifetime psychotic disorder. At baseline, 1119 patients, 1059 first-degree siblings and 586 healthy controls were recruited. Details on the GROUP project structure, recruitment of participants, sample size estimation, data collection and ethical approval have been published elsewhere [27].

2.2. Measurements

Data were collected at baseline and after three years and six years of follow-up. DSM-4 [28] was used to diagnose schizophrenia spectrum disorder. Positive and negative symptoms in patients were assessed at baseline and after three years and six years using the Positive and Negative Syndrome Scale (PANSS) [19,20]. Similarly, positive and negative schizotypy (i.e., subclinical symptoms) in siblings and healthy controls were assessed with the Structured Interview for Schizotypy-revised (SIS-R) [19,20]. Positive and negative subscales' mean or sum score was used to quantify the severity of symptoms. Cognitive function was assessed in patients, siblings and controls at baseline and after three years and six years using the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) consensus cognitive battery test [21–23]. Specifically, we used the Word Learning Task (WLT) (i.e., to measure memory), Continuous Performance Test-HQ (CPT-HQ) (i.e., to measure executive function) and Wechsler Adult Intelligence Scale (WAIS-III) (i.e., to measure IQ). Subdomain and global cognitive function were evaluated by creating a composite score using the Z-score [22,23] or principal component analysis methods [21]. Social function was measured after three years and six years by using the Social Functioning Scale (SFS), whereas quality of life was measured at baseline and after three years and six years with the brief version of the World Health Organization Quality of Life assessment (WHOQOL-BREF) [24]. A multidimensional social inclusion composite score was created from three years measurement of the SFS and WHOQOL-BREF using the Z-score method [24]. Premorbid adjustment was retrospectively measured using the premorbid adjustment sale (PAS) in three life periods: childhood (<12 years), early adolescence (12–16 years) and late adolescence (16–19 years) [25,26]. All information was gathered from parents or siblings and a mean score was created to quantify overall premorbid functioning [25,26]. Sociodemographic, genetic and clinical data were also collected. DNA was extracted from peripheral blood lymphocytes in a 20 mL stored blood sample. Genotyping, quality control and imputation were carried out following the standard procedure [19,21]. The polygenic risk score for schizophrenia (PRS$_{SCZ}$) was calculated using Psychiatric Genomics Consortium (PGC) schizophrenia genome-wide association study (GWAS) summary statistics, where a higher PRS$_{SCZ}$ indicates a higher genetic vulnerability to developing schizophrenia [21].

2.3. Data Analyses

Identification of latent subtypes: Person-centred latent class models were used to identify subgroups of individuals with similar phenotypes, which we call latent subtypes throughout this article. In five studies [19–22,26], we applied group-based trajectory modelling (GBTM) to identify latent subtypes and evaluate their longitudinal course throughout the six-year follow-up, whereas we applied K-means clustering in three studies [23–25] to identify latent subtypes at one time point (e.g., baseline data, or three-year follow-up). The analyses were carried out using SAS and R software.

Predictors of latent subtypes: In three studies [19,21,22], mixed-effects regression models were applied to investigate predictors of latent subtypes. The other studies [20,23–25] used analysis of variance (ANOVA) or the Kruskal–Wallis test and chi-square test to investigate continuous and categorical variables that distinguished latent subtypes, respectively. The analyses were carried out using SAS and R software.

3. Results

3.1. Latent Subtypes

Positive and negative symptoms: In patients, we identified three latent subtypes of positive and negative symptoms [19] (Figure 1A). Additionally, using social amotivation and expressive deficit subphenotypes of negative symptoms, we identified four latent subtypes in patients [20]. In siblings, we consistently distinguished four latent subtypes of positive and negative schizotypy (Figure 1B), while three positive and negative schizotypy latent subtypes were identified in controls (Figure 1C) [19].
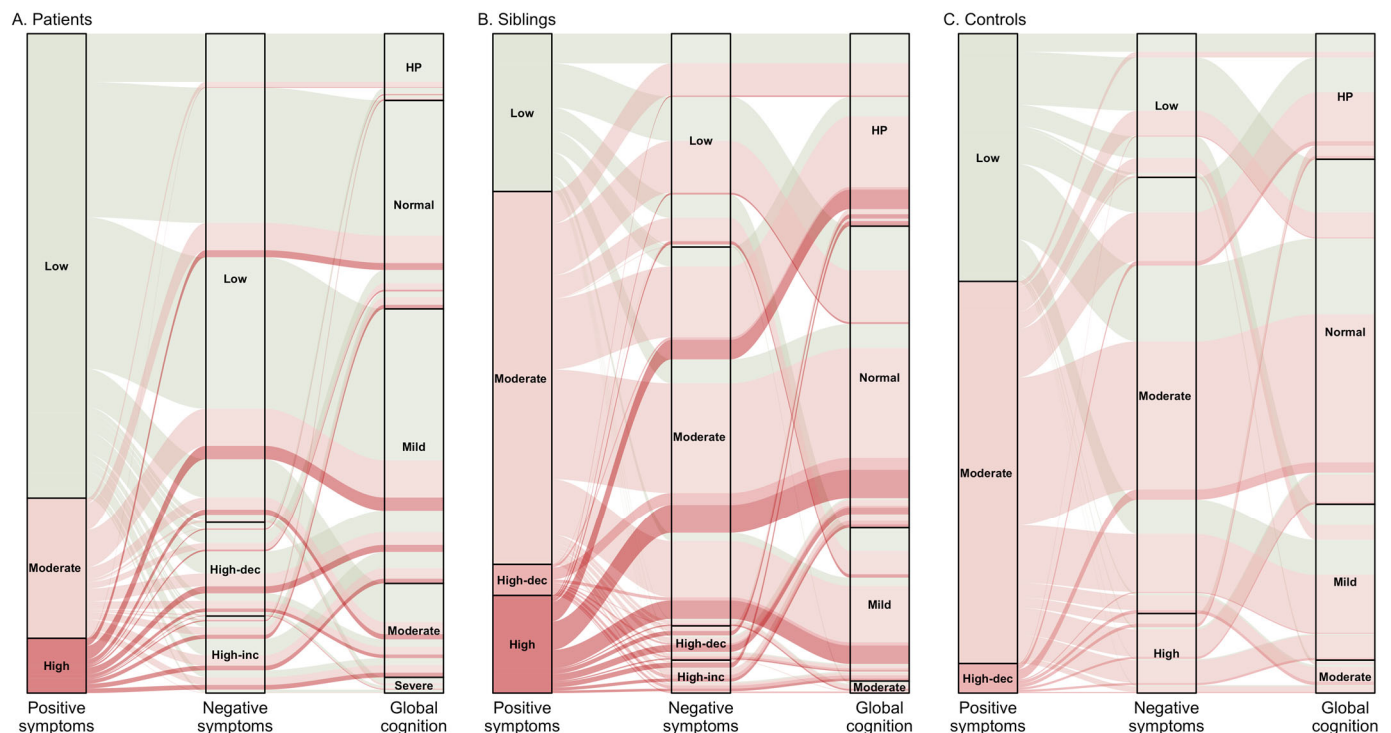
Figure 1. Spectrum of positive, negative and cognitive symptom subtypes in patients (**A**), siblings (**B**) and controls (**C**). High-dec: high-decreased; High-inc: high-increased; HP: high-performing. Cognitive subtypes coded from best to worse as high-performing, normal, mildly impaired, moderately impaired, severely impaired and very severely impaired.

Cognitive functions: We identified five and four cognitive subtypes in patients and their siblings, respectively (Figure 1A,B) [21,22]. In another study using cross-sectional data, we identified three cognitive subtypes in siblings [23]. In controls, we identified four cognitive subtypes (Figure 1C) [21]. After combing all samples together, we unravelled six latent

cognitive subtypes (Figure 2) [21]. Furthermore, we examined the latent subtypes of three cognitive subdomains—verbal learning and memory, sustained attention and vigilance, and global cognitive function (IQ) [29]. In patients, we identified five latent subtypes of global cognitive function (IQ), and three subtypes of verbal learning and memory, and sustained attention and vigilance. In siblings and controls, five, four and three subtypes of global cognitive function (IQ), verbal learning and memory, and sustained attention and vigilance were identified, respectively. In all combined samples (Figure 2), we identified five subtypes of verbal learning and memory, and global cognitive function (IQ) domains, whereas four subtypes of sustained attention and vigilance were identified.
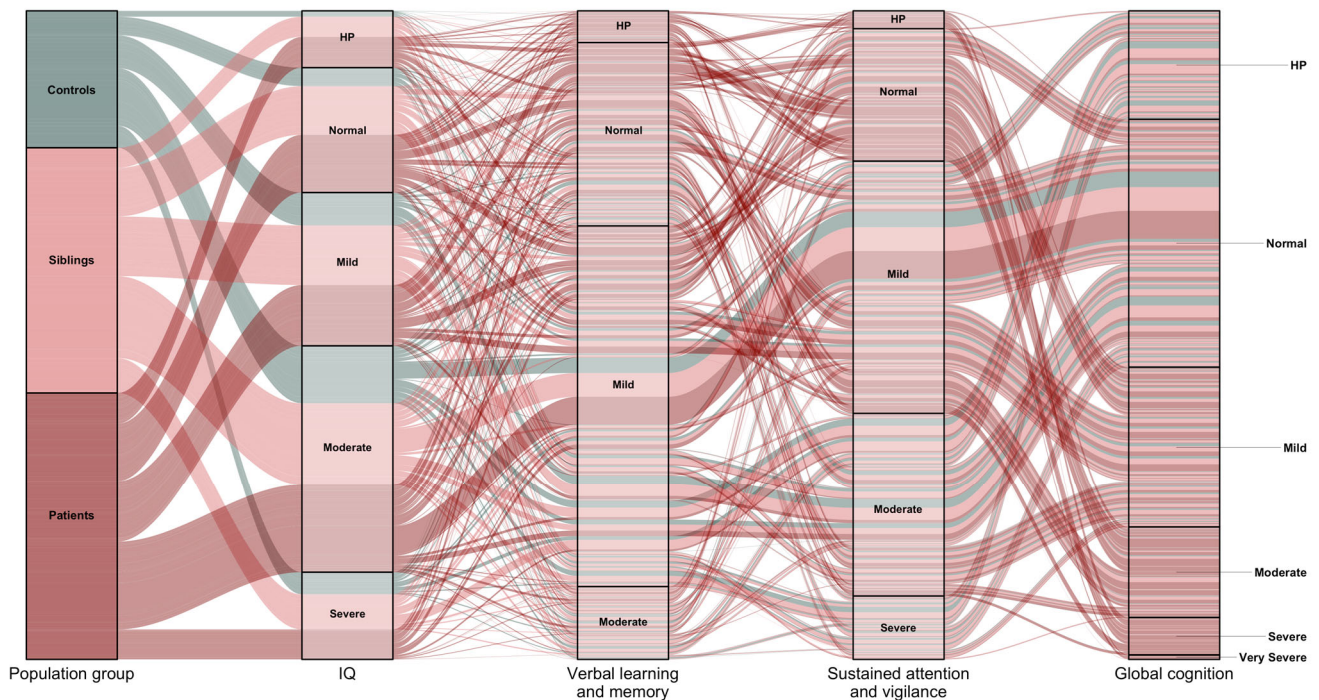
Figure 2. Spectrum of cognitive symptom subtypes in whole combined sample. HP: high-performing. Cognitive subtypes coded from best to worse as high-performing, normal, mildly impaired, moderately impaired, severely impaired and very severely impaired.

Psychosocial functions: In patients, we identified five latent subtypes of multidimensional social inclusion [24]. In addition, we revealed six latent subtypes of premorbid adjustment in patients [25,26].

3.2. Longitudinal Courses and Profiles of Subtypes

Positive and negative symptoms: Positive and negative symptoms had stable, deteriorating, relapsing and ameliorating courses over time in patients, siblings and controls [19,20].

Cognitive functions: Cognitive subtypes had stable course during the six-year period in patients, siblings, controls and combined samples [21,22]. Additionally, cluster analyses showed "normal", "mixed" and "impaired" cognitive profiles in siblings [23].

Psychosocial functions: Premorbid adjustment had "normal, slow decrease", "normal, rapid decrease", "mild, rapid decrease", "mild, slow decrease", "moderate, slow decrease" and "severe, slow decrease" longitudinal courses over the six-year period [26]. Another study showed "normal", "social intermediate", "academic decline", "overall decline", "overall intermediate" and "overall impaired" premorbid adjustment profiles in patients [25]. Multidimensional social inclusion had "very low (social functioning)/very low (quality of life)", "low/low", "high/low", "medium/high" and "high/high" profiles [24].

3.3. Predictors of Latent Subtypes

Positive and negative symptoms: In patients, baseline positive symptom severity predicted latent subtypes of positive symptoms, while baseline negative symptom severity and quality of life predicted latent subtypes of negative symptoms [19]. In siblings, psychotic-like experiences (i.e., depressive symptoms) and quality of life predicted positive schizotypy subtypes, whereas premorbid adjustment, psychotic-like experiences (i.e., positive symptoms) and quality of life predicted negative schizotypy subtypes [19]. In controls, quality of life predicted positive schizotypy subtypes, while premorbid adjustment and quality of life predicted negative schizotypy subtypes.

Cognitive functions: Premorbid IQ, age and premorbid adjustment predicted cognitive latent subtypes in patients [21]. Age and poor functional outcomes predicted cognitive subtypes in siblings [21,23]. Interestingly, the cognitive symptoms of siblings were predicted by the cognitive subtypes of patients [21]. Age, premorbid IQ, premorbid adjustment and positive schizotypy predicted cognitive subtypes in controls [21]. PRS_{SCZ}, age, gender, ethnicity, educational status, premorbid IQ and premorbid adjustment predicted cognitive subtypes in combined samples [22].

Psychosocial functions: PRS_{SCZ}, positive, negative and depressive symptoms, premorbid social functioning, number of met needs and satisfaction predicted multidimensional social inclusion subtypes [24].

4. Discussion

Through deep clinical phenotyping of SSD using the Genetic Risk and Outcome of Psychosis (GROUP) cohort dataset, we identified three to four latent subtypes of positive and negative symptoms [19,20] and four to six latent cognitive subtypes [21–23] in patients, siblings and controls. In our previous systematic review of worldwide longitudinal and cross-sectional data-driven studies, we observed two to five latent subtypes of positive, negative and cognitive symptoms in patients and controls [30]. In line with previous studies [31–33], we also identified five latent subtypes of psychosocial function—multidimensional social inclusion and premorbid adjustment—in patients [24–26]. Our findings support the DSM-5 dimensional approach for characterizing disorders based on a collection of symptoms [3]. Taken together, these findings demonstrate the inherent heterogeneity of SSD. Data-driven approaches have been shown to have the ability to explore heterogeneity and biological drivers of SSD and other psychotic disorders in the last decade [5]. Furthermore, data-driven approaches could contribute to the precise identification of high-risk population groups, optimal selection of interventions targeting subgroups of patients with similar phenotypes and evaluation of patients' prognosis [34]. Consequently, they could help foster precision psychiatry by tackling diagnostic and treatment selection challenges pertaining to heterogeneity.

The identified latent subtypes in the GROUP studies [19–26] exhibited stable, deteriorating, improving and relapsing clinical courses over time with mixed phenotypes. We observed similar patterns of phenotypes in studies (particularly in patients and controls) included in our previous systematic review [30]. This implies the relevance of considering dynamic phenotypical changes over time, and therapies need to be seen as a process of change [35–37]. Interestingly, cognitive subtypes showed a stable course during the six-year follow-up [21,22]. This finding supports that cognitive deficits continue to be a valuable endophenotype of interest in the search for specific genetic factors related to schizophrenia [38,39]. The different symptom subtypes and their developmental courses appearing in both subclinical and clinical phases may elucidate the disease mechanisms before the disease onset and may be helpful for predicting transition to psychosis in unaffected individuals, which provides ample opportunities for early recognition, delaying onset and the treatment of subtle psychotic symptoms [40].

Even though up to 58 factors have been suggested to predict subtypes, only limited factors predicted latent subtypes in our studies [30]. This is possibly because we used multivariable models, whereas most of the other studies used univariable models or only

pair-wise comparison was carried out [30]. Interestingly, we observed an overlap of factors that predict positive, negative and cognitive symptoms and psychosocial function subtypes. This suggests that these phenotypes have shared pathophysiology and similar clinical courses. Similarly, recently emerging evidence [41–43] showed a temporal association between positive, negative and cognitive symptoms and identified common predictive factors [44].

PRS_{SCZ} showed better association with cognitive subtypes than positive and negative symptom subtypes [21]. This suggests that cognition may have higher genetic susceptibility than positive and negative symptoms. Cognitive subtypes can be a suitable endophenotype of SSD [45] given the stable longitudinal course and its association with PRS_{SCZ}. Endophenotype has been suggested to have a simpler genetic architecture than clinical syndromes and it can provide better signal-to-noise ratios and greater statistical power in research, which can be detectable in smaller sample sizes [45]. Furthermore, the strong predictive power of PRS_{SCZ} in the combined sample demonstrates the relevance of a large sample size and partly demonstrates the possibility of translating the constructed PRS_{SCZ} in research settings into actual practices to disentangle clinically relevant patient subtypes [46,47]. Broadly, the use of PRS is promising for early detection, the initiation of prevention strategies and the prediction of treatment outcomes, although studies that investigated the role of PRS_{SCZ} to predict latent subtypes are limited [30] and challenges exist in translational psychiatry research.

Heterogeneity in the traditional classification of mental disorders, including schizophrenia spectrum disorders (SSD), is the main challenge in research and clinical practice [48,49]. As such, key scientific values embodied in this article are highlighted as follows. Firstly, we combined data-driven, polygenic scoring and subphenotyping approaches to evaluate the profiles and longitudinal courses of SSD phenotypes by incorporating not only clinical but also psychosocial outcomes in patients and siblings. This is a unique aspect that has not been extensively explored in previous clinical studies. To best of our knowledge, there is no cohort study that comprehensively investigated positive and negative symptoms, cognitive impairments and social outcomes and compared findings across patients, siblings and controls. Most previous studies focused on positive and negative symptoms in patients, followed by cognitive impairment [30]. Studies that consistently examined the role of polygenic risk score to predict latent subtypes in positive and negative symptoms are very limited, let alone in social outcomes. Secondly, we depicted the presence of high heterogeneity and tried to show the full picture of SSD phenotypes, which is not addressed in the previous literature. Thirdly, we conducted comprehensive data analysis using various complementary statistical modelling techniques to obtain insight into the prominent predictors of latent subtypes. The combination of using data-driven approaches and the inclusion of PRS and exposomes enhances our understanding of SSD and its trajectory over time. Lastly, the novel findings presented in this article—particularly in the sibling population and psychosocial functioning phenotypes—can substantially contribute to the existing body of scientific knowledge and open new avenues for future investigations and potential clinical applications.

The findings presented in this article also have clinical relevances and will contribute to a broader understanding of the clinical heterogeneity and disease course in SSD in two ways. First, data-driven methods can identify subgroups of SSD patients based on their phenotypes, which may ultimately help refine psychiatric nosology [49]. Additionally, this helps to reduce the heterogeneity for a more targeted assessment of clinical and biological predictors of these phenotypes in SSD. The large number of subtypes observed in studies also alerts clinicians to not use generic approaches to treat patients and consider siblings or family members during treatment. Second, the identification of specific biomarkers and clinical factors that are associated with specific latent subtypes will allow for the tailoring of treatments to alleviate positive and negative symptoms and improve cognitive performance and psychosocial functioning in SSD. We identified the predictors that can be used as a marker for clinicians for precisely identifying high-risk groups or patients

with similar phenotypes. This could minimize the misdiagnosis that commonly seen in schizophrenia [13]. Overall, these findings may have the potential to impact patient care and early preventive and treatment strategies, leading to improved patient outcomes in the long term.

5. Strengths and Limitations

This article presented evidence on in-depth phenotyping of SSD targeting for multiple phenotypes in patients, siblings and controls. We identified latent subtypes using cross-sectional and longitudinal data. The longitudinal data also helped us to examine the long-term clinical course of phenotypes [13]. We also investigated various factors that predict latent subtypes and could be helpful for understanding disease progression. On the other hand, we presented findings based on a single cohort sample, which may limit the generalizability of our findings. However, our findings are comparable to results from a review of previous data-driven studies [30].

6. Future Research Directions

Data complexity, ambiguous subtyping results and nomenclature, and the complex relationship between phenotypes are still bottlenecks to fully harnessing the advantages of data-driven methods [12,48,50]. In the attempts to carry out in-depth phenotyping of SSD and the realization of precision psychiatric care, future collaborative research is warranted.

First, the identified latent subtypes and their clinical courses need to be validated in independent cohorts to ensure the generalizability of results. Additionally, more longitudinal studies are needed to evaluate the long-term course of cognitive and functional outcomes in patients and siblings.

Second, subphenotyping targeting specific cognitive functions (e.g., memory, attention), positive symptoms (e.g., hallucination, delusion, disorganized thought), negative symptoms (e.g., asociality, avolition, affective flattening) and psychosocial functions (e.g., social engagement, housing, employment) in patients, siblings and controls using data-driven approaches may provide additional insights for deeper characterization of SSD. Endophenotype studies should also continue from interdisciplinary domains to address knowledge gaps and to deliver clinically applicable insights [45]. Additionally, more studies on siblings of patients with SSD are needed, as they are frequently affected due to shared genetic and non-genetic risk factors. For example, only 3 out of 53 worldwide data-driven studies were conducted on siblings [30]. Further data-driven studies on psychosocial function in patients and siblings are warranted.

Third, the use of growth mixture models or a dynamic mixture of expert models [51,52] may provide robust results, as these models allow the incorporation of both time-varying predictors and baseline predictors in the clustering of phenotypes and capture the dynamics better over time.

Fourth, as efforts have been devoted to unravelling the pathophysiology of SSD, it can also be beneficial to further examine the interrelatedness and distinctiveness of positive and negative symptoms, cognitive impairments and psychosocial functions [42] using causal discovery models [53,54].

Fifth, investigating the association of PRS_{SCZ} with phenotypes, endophenotypes, biological substrates and metabolites in large multinational consortia could help the identification of symptoms that have a stronger genetic liability and understanding of underlying causal models specific to clinical subgroups or symptom dimensions [55,56]. Performing GWAS of the trajectory of SSD phenotypes [57] and using the PRS of these phenotypes could be better than PRS_{SCZ} to deeply characterize latent subtypes [58,59]. Furthermore, calculating a polyomic risk score by integrating phenomics, genomics, epigenomics, proteomics, transcriptomics and pharmacogenomics may increase the performance of risk prediction, patients' stratification for interventions and the efficacy of treatments [60,61].

Sixth, future studies may focus on enriching data-driven models based on clearly defined research questions and developing supplementary techniques for interpreting

results and successfully translating findings into clinical practice [12,48]. It is also important to develop a pipeline composed of data collection, pre-processing and data modelling to identify subtypes with different profiles or disease progressions and risks of disease complications [18].

Seventh, future studies can be benefited by including inflammatory biomarkers, cardiometabolic data and lifestyle factors in patients and siblings, which could be critical for understanding the pathophysiology of somatic comorbidities that occur before or concomitantly with SSD [41]. In our previous study, we observed increased HbA1c levels in patients with SSD [62]. Using insulin sensitivity measures (euglycemic clamp), energy measurements and body composition could be also helpful for a more in-depth understanding of cardiometabolic co-morbidities that occur before and with SSD. Furthermore, extensive investigation of the role of antipsychotics and adherence to treatment could be helpful for precisely determining the clinical course of latent subtypes.

Eighth, applying Bayesian statistical models could provide strong evidence when the sample size is small, usability of omics data is limited, and follow-up duration is short compared to conventional statistical models. Bayesian inference is known to have an advantage given that the results are not dependent on the assumption of a large sample size, and it creates an opportunity to incorporate expert knowledge into the data-driven approach [63]. Therefore, the investigation of latent subtypes and their predictors using Bayesian latent class models can be an additional future research area [64,65].

Finally, recent advances in the theory and applicability of machine learning methods constitute an important area of future research due to the complexity of the data structure. Thus, hybrid methods that combine both supervised and unsupervised methods can identify subpopulations, contribute to comparing identified subpopulations and may provide superior results for the robust selection of predictors, which may ultimately help refine psychiatric nosology [49,66].

7. Conclusions

SSD phenotypes are highly heterogenous and show mixed profiles and longitudinal course among patients, siblings and controls. Three to four latent subtypes of positive and negative symptoms and four to six latent cognitive subtypes were identified in patients, siblings and controls. Five latent subtypes of psychosocial function—multidimensional social inclusion and premorbid adjustment—were identified in patients. We depicted that the identified subtypes had mixed profiles and stable, deteriorating, relapsing and ameliorating courses during the six-year follow-up period. Baseline positive and negative symptoms, premorbid adjustment, psychotic-like experiences, health-related quality of life and PRS_{SCZ} were found to be the strong predictors of the identified subtypes.

In the current era with limited objective criteria to diagnose SSD, the application of data-driven approaches could be relevant for clinicians and researchers for in-depth phenotyping of SSD by compiling all symptoms' domains across patients and healthy individuals. The identification of latent subtypes and their clinical courses may help to design randomized control trials and implement targeted evidence-based interventions with maximized treatment efficacy and minimized costs and side effects.

Author Contributions: T.D.H. and B.Z.A. conceived and design the study. T.D.H. and J.H. drafted and revised the manuscript. B.Z.A. and R.B. are the principal investigators and supervisors of GROUP cohort studies. They also contributed to the critical revision of the manuscript. E.J.L. and N.B. contributed the critical revision and methodological insights of the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Owen, M.J.; Sawa, A.; Mortensen, P.B. Schizophrenia. *Lancet* **2016**, *388*, 86–97. [CrossRef] [PubMed]
2. Charlson, F.J.; Ferrari, A.J.; Santomauro, D.F.; Diminic, S.; Stockings, E.; Scott, J.G.; McGrath, J.J.; Whiteford, H. Global Epidemiology and Burden of Schizophrenia: Findings from the Global Burden of Disease Study 2016. *Schizophr. Bull.* **2018**, *44*, 1195–1203. [CrossRef] [PubMed]
3. Association, A.P. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*; American Psychiatric Pub: Arlington, VA, USA, 2013.
4. McCutcheon, R.; Marques, T.R.; Howes, O.D. Schizophrenia—An Overview. *JAMA Psychiatry* **2020**, *77*, 201–210. [CrossRef] [PubMed]
5. Correll, C.U.; Schooler, N.R. Negative Symptoms in Schizophrenia: A Review and Clinical Guide for Recognition, Assessment, and Treatment. *Neuropsychiatr. Dis. Treat.* **2020**, *16*, 519–534. [CrossRef]
6. Zarogianni, E.; Storkey, A.J.; Johnstone, E.C.; Owens, D.G.; Lawrie, S.M. Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. *Schizophr. Res.* **2017**, *181*, 6–12. [CrossRef]
7. Keefe, R.S.E.; Fenton, W.S. How Should DSM-V Criteria for Schizophrenia Include Cognitive Impairment? *Schizophr. Bull.* **2007**, *33*, 912–920. [CrossRef]
8. Walker, A.E.; Spring, J.D.; Travis, M.J. Addressing Cognitive Deficits in Schizophrenia: Toward a Neurobiologically Informed Approach. *Biol. Psychiatry* **2017**, *81*, e1–e3. [CrossRef]
9. Lencz, T.; Knowles, E.; Davies, G.; Guha, S.; Liewald, D.C.; Starr, J.M.; Djurovic, S.; Melle, I.; Sundet, K.; Christoforou, A.; et al. Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: A report from the Cognitive Genomics consorTium (COGENT). *Mol. Psychiatry* **2013**, *19*, 168–174. [CrossRef]
10. Blackman, R.K.; Dickinson, D.; Eisenberg, D.P.; Gregory, M.D.; Apud, J.A.; Berman, K.F. Antipsychotic medication-mediated cognitive change in schizophrenia and polygenic score for cognitive ability. *Schizophr. Res. Cogn.* **2021**, *27*, 100223. [CrossRef]
11. Cooper, A.M.; Michels, R. Diagnostic and statistical manual of mental disorders, revised (DSM-III-R). *Am. J. Psychiatry* **1988**, *145*, 1300–1301. [CrossRef]
12. Yung, A.R.; Phillips, L.J.; Yuen, H.P.; McGorry, P.D. Risk factors for psychosis in an ultra high-risk group: Psychopathology and clinical features. *Schizophr. Res.* **2004**, *67*, 131–142. [CrossRef] [PubMed]
13. Lopez-Castroman, J.; Leiva-Murillo, J.M.; Cegla-Schvartzman, F.; Blasco-Fontecilla, H.; Garcia-Nieto, R.; Artes-Rodriguez, A.; Morant-Ginestar, C.; Courtet, P.; Blanco, C.; Aroca, F.; et al. Onset of schizophrenia diagnoses in a large clinical cohort. *Sci. Rep.* **2019**, *9*, 9865. [CrossRef] [PubMed]
14. Korth, C.; Fangerau, H. Blood tests to diagnose schizophrenia: Self-imposed limits in psychiatry. *Lancet Psychiatry* **2020**, *7*, 911–914. [CrossRef] [PubMed]
15. Stuhec, M. Antipsychotic treatment in elderly patients on polypharmacy with schizophrenia. *Curr. Opin. Psychiatry* **2022**, *35*, 332–337. [CrossRef]
16. Muthen, B.; Muthen, L.K. Integrating Person-Centered and Variable-Centered Analyses: Growth Mixture Modeling with Latent Trajectory Classes. *Alcohol. Clin. Exp. Res.* **2000**, *24*, 882–891. [CrossRef]
17. Delude, C.M. Deep phenotyping: The details of disease. *Nature* **2015**, *527*, S14–S15. [CrossRef]
18. Robinson, P.N. Deep phenotyping for precision medicine. *Hum. Mutat.* **2012**, *33*, 777–780. [CrossRef] [PubMed]
19. Habtewold, T.D.; Tiles-Sar, N.; Liemburg, E.J.; Sandhu, A.K.; Islam, M.A.; Boezen, H.M.; Bruggeman, R.; Alizadeh, B.Z. Six-year trajectories and associated factors of positive and negative symptoms in schizophrenia patients, siblings, and controls: The GROUP study. *Res. Sq.* **2023**. [CrossRef]
20. Stiekema, A.P.; Islam, A.; Liemburg, E.J.; Castelein, S.; Heuvel, E.R.V.D.; van Weeghel, J.; Aleman, A.; Bruggeman, R.; van der Meer, L. Long-term course of negative symptom subdomains and relationship with outcome in patients with a psychotic disorder. *Schizophr. Res.* **2018**, *193*, 173–181. [CrossRef]

21. Habtewold, T.D.; Liemburg, E.J.; Islam, A.; de Zwarte, S.M.; Boezen, H.M.; Bruggeman, R.; Alizadeh, B.Z.; Luykx, J.J.; Rutten, B.P.; van Winkel, R.; et al. Association of schizophrenia polygenic risk score with data-driven cognitive subtypes: A six-year longitudinal study in patients, siblings and controls. *Schizophr. Res.* **2020**, *223*, 135–147. [CrossRef]

22. Islam, A.; Habtewold, T.D.; van Es, F.D.; Quee, P.J.; Heuvel, E.R.V.D.; Alizadeh, B.Z.; Bruggeman, R.; GROUP Investigators. Long-term cognitive trajectories and heterogeneity in patients with schizophrenia and their unaffected siblings. *Acta Psychiatr. Scand.* **2018**, *138*, 591–604. [CrossRef]

23. Quee, P.J.; Alizadeh, B.Z.; Aleman, A.; Heuvel, E.R.V.D.; GROUP Investigators. Cognitive subtypes in non-affected siblings of schizophrenia patients: Characteristics and profile congruency with affected family members. *Psychol. Med.* **2014**, *44*, 395–405. [CrossRef] [PubMed]

24. Hao, J.; Tiles-Sar, N.; Liemburg, E.J.; Habtewold, T.D.; Bruggeman, R.; van der Meer, L.; Alizadeh, B.Z. Multidimensional social inclusion and its prediction in schizophrenia spectrum disorder. *Res. Sq.* **2023**.

25. Quee, P.J.; GROUP Investigators; Meijer, J.H.; Islam, A.; Aleman, A.; Alizadeh, B.Z.; Meijer, C.J.; Heuvel, E.R.V.D. Premorbid adjustment profiles in psychosis and the role of familial factors. *J. Abnorm. Psychol.* **2014**, *123*, 578–587. [CrossRef] [PubMed]

26. Tiles-Sar, N.; Habtewold, T.D.; Liemburg, E.J.; van der Meer, L.; Bruggeman, R.; Alizadeh, B.Z.; van Amelsvoort, T.; Bartels-Velthuis, A.; de Haan, L.; Schirmbeck, F.; et al. Understanding Lifelong Factors and Prediction Models of Social Functioning After Psychosis Onset Using the Large-Scale GROUP Cohort Study. *Schizophr. Bull.* **2023**, sbad046. [CrossRef]

27. Korver, N.; Quee, P.J.; Boos, H.B.M.; Simons, C.J.P.; de Haan, L.; GROUP Investigators. Genetic Risk and Outcome of Psychosis (GROUP), a multi site longitudinal cohort study focused on gene-environment interaction: Objectives, sample characteristics, recruitment and assessment methods. *Int. J. Methods Psychiatr. Res.* **2012**, *21*, 205–221. [CrossRef]

28. Bell, C.C. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA* **1994**, *272*, 828–829. [CrossRef]

29. Habtewold, T. Data-Driven Subphenotypic Dissection of the Clinical Heterogeneity of Schizophrenia Spectrum Disorders. Ph.D. Thesis, University of Groningen, Groningen, The Netherlands, 2021.

30. Habtewold, T.D.; Rodijk, L.H.; Liemburg, E.J.; Sidorenkov, G.; Boezen, H.M.; Bruggeman, R.; Alizadeh, B.Z. A systematic review and narrative synthesis of data-driven studies in schizophrenia symptoms and cognitive deficits. *Transl. Psychiatry* **2020**, *10*, 224. [CrossRef]

31. Hall, M.-H.; Holton, K.M.; Öngür, D.; Montrose, D.; Keshavan, M.S. Longitudinal trajectory of early functional recovery in patients with first episode psychosis. *Schizophr. Res.* **2019**, *209*, 234–244. [CrossRef]

32. Chang, W.C.; Chu, A.O.K.; Kwong, V.W.Y.; Wong, C.S.M.; Hui, C.L.; Chan, K.W.; Lee, H.M.E.; Chen, E. Patterns and predictors of trajectories for social and occupational functioning in patients presenting with first-episode non-affective psychosis: A three-year follow-up study. *Schizophr. Res.* **2018**, *197*, 131–137. [CrossRef]

33. Kam, C.T.K.; Chang, W.C.; Kwong, V.W.Y.; Lau, E.S.K.; Chan, G.H.K.; Jim, O.T.T.; Hui, C.L.M.; Chan, S.K.W.; Lee, E.H.M.; Chen, E.Y.H. Patterns and predictors of trajectories for subjective quality of life in patients with early psychosis: Three-year follow-up of the randomized controlled trial on extended early intervention. *Aust. N. Z. J. Psychiatry* **2021**, *55*, 983–992. [CrossRef] [PubMed]

34. Mori, M.; Krumholz, H.M.; Allore, H.G. Using Latent Class Analysis to Identify Hidden Clinical Phenotypes. In *JAMA Guide to Statistics and Methods*; Livingston, E.H., Lewis, R.J., Eds.; McGraw-Hill Education: New York, NY, USA, 2019. [CrossRef]

35. Hayes, A.M.; Feldman, G. Clarifying the construct of mindfulness in the context of emotion regulation and the process of change in therapy. *Clin. Psychol. Sci. Pract.* **2004**, *11*, 255–262. [CrossRef]

36. Hayes, S.C.; Hofmann, S.G. The third wave of cognitive behavioral therapy and the rise of process-based care. *World Psychiatry* **2017**, *16*, 245–246. [CrossRef]

37. Zanardi, R.; Prestifilippo, D.; Fabbri, C.; Colombo, C.; Maron, E.; Serretti, A. Precision psychiatry in clinical practice. *Int. J. Psychiatry Clin. Pract.* **2020**, *25*, 19–27. [CrossRef] [PubMed]

38. Snitz, B.; MacDonald, A.W., III; Carter, C.S. Cognitive Deficits in Unaffected First-Degree Relatives of Schizophrenia Patients: A Meta-analytic Review of Putative Endophenotypes. *Schizophr. Bull.* **2005**, *32*, 179–194. [CrossRef]

39. Greenwood, T.A.; Lazzeroni, L.C.; Calkins, M.E.; Freedman, R.; Green, M.F.; Gur, R.E.; Gur, R.C.; Light, G.A.; Nuechterlein, K.H.; Olincy, A.; et al. Genetic assessment of additional endophenotypes from the Consortium on the Genetics of Schizophrenia Family Study. *Schizophr. Res.* **2016**, *170*, 30–40. [CrossRef]

40. van Os, J.; Guloksuz, S. A critique of the "ultra-high risk" and "transition" paradigm. *World Psychiatry* **2017**, *16*, 200–206. [CrossRef]

41. Chen, Y.-L.; Pan, C.-H.; Chang, C.-K.; Chen, P.-H.; Chang, H.-M.; Tai, M.-H.; Su, S.-S.; Tsai, S.-Y.; Chen, C.-C.; Kuo, C.-J. Physical Illnesses Before Diagnosed as Schizophrenia: A Nationwide Case-Control Study. *Schizophr. Bull.* **2020**, *46*, 785–794. [CrossRef]

42. Eack, S.M.; Keshavan, M.S. Cognition, negative symptoms, and functional outcome in psychosis. *Schizophr. Res.* **2020**, *224*, 22–23. [CrossRef]

43. Harvey, P.D.; Koren, D.; Reichenberg, A.; Bowie, C.R. Negative Symptoms and Cognitive Deficits: What Is the Nature of Their Relationship? *Schizophr. Bull.* **2006**, *32*, 250–258. [CrossRef]

44. Carrà, G.; Crocamo, C.; Angermeyer, M.; Brugha, T.; Toumi, M.; Bebbington, P. Positive and negative symptoms in schizophrenia: A longitudinal analysis using latent variable structural equation modelling. *Schizophr. Res.* **2019**, *204*, 58–64. [CrossRef] [PubMed]

45. Roffman, J.L. Endophenotype Research in Psychiatry—The Grasshopper Grows Up. *JAMA Psychiatry* **2019**, *76*, 1230. [CrossRef] [PubMed]

46. Binder, E.B. Polygenic risk scores in schizophrenia: Ready for the real world? *Am. Psychiatr. Assoc.* **2019**, *174*, 783–784. [CrossRef]

47. Zheutlin, A.B.; Dennis, J.; Linnér, R.K.; Moscati, A.; Restrepo, N.; Straub, P.; Ruderfer, D.; Castro, V.M.; Chen, C.-Y.; Ge, T.; et al. Penetrance and Pleiotropy of Polygenic Risk Scores for Schizophrenia in 106,160 Patients Across Four Health Care Systems. *Am. J. Psychiatry* **2019**, *176*, 846–855. [CrossRef] [PubMed]

48. Allsopp, K.; Read, J.; Corcoran, R.; Kinderman, P. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Res.* **2019**, *279*, 15–22. [CrossRef]

49. Feczko, E.; Miranda-Dominguez, O.; Marr, M.; Graham, A.M.; Nigg, J.T.; Fair, D.A. The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends Cogn. Sci.* **2019**, *23*, 584–601. [CrossRef]

50. Twisk, J.; Hoekstra, T. Classifying developmental trajectories over time should be done with great caution: A comparison between methods. *J. Clin. Epidemiol.* **2012**, *65*, 1078–1087. [CrossRef]

51. Quiroz, M.; Villani, M. Dynamic mixture-of-experts models for longitudinal and discrete-time survival data. *Riksbank Res. Pap. Ser.* **2013**, *99*, 39. [CrossRef]

52. Ram, N.; Grimm, K.J. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *Int. J. Behav. Dev.* **2009**, *33*, 565–576. [CrossRef]

53. Glymour, C.; Zhang, K.; Spirtes, P. Review of Causal Discovery Methods Based on Graphical Models. *Front. Genet.* **2019**, *10*, 524. [CrossRef]

54. Rohrer, J.M. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Adv. Methods Prac. Psychol. Sci.* **2018**, *1*, 27–42. [CrossRef]

55. Fanous, A.H.; Kendler, K.S. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: Searching for a framework. *Mol. Psychiatry* **2005**, *10*, 6–13. [CrossRef] [PubMed]

56. Jones, H.J.; Heron, J.; Hammerton, G.; Stochl, J.; Jones, P.B.; Cannon, M.; Smith, G.D.; Holmans, P.; Lewis, G.; Linden, D.E.J.; et al. Investigating the genetic architecture of general and specific psychopathology in adolescence. *Transl. Psychiatry* **2018**, *8*, 145. [CrossRef] [PubMed]

57. Ko, S.; German, C.A.; Jensen, A.; Shen, J.; Wang, A.; Mehrotra, D.V.; Sun, Y.V.; Sinsheimer, J.S.; Zhou, H.; Zhou, J.J. GWAS of longitudinal trajectories at biobank scale. *Am. J. Hum. Genet.* **2022**, *109*, 433–445. [CrossRef] [PubMed]

58. Bigdeli, T.B.; Peterson, R.E.; Ripke, S.; Bacanu, S.-A.; Amdur, R.L.; Gejman, P.V.; Levinson, D.F.; Riley, B.R.; St. Clair, D.; Rietschel, M.; et al. Genome-wide association study of clinical features in the schizophrenia psychiatric genomics consortium: Confirmation of polygenic effect on negative symptoms. *bioRxiv* **2017**. [CrossRef]

59. Jonas, K.G.; Lencz, T.; Li, K.; Malhotra, A.K.; Perlman, G.; Fochtmann, L.J.; Bromet, E.J.; Kotov, R. Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. *Transl. Psychiatry* **2019**, *9*, 300. [CrossRef] [PubMed]

60. Arehart, C.H.; Sterrett, J.D.; Garris, R.L.; Quispe-Pilco, R.E.; Gignoux, C.R.; Evans, L.M.; Stanislawski, M.A. Poly-omic risk scores predict inflammatory bowel disease diagnosis. *bioRxiv* **2022**, 508056.

61. Wheeler, H.E.; Aquino-Michaels, K.; Gamazon, E.R.; Trubetskoy, V.V.; Dolan, M.E.; Huang, R.S.; Cox, N.J.; Im, H.K. Poly-Omic Prediction of Complex Traits: OmicKriging. *Genet. Epidemiol.* **2014**, *38*, 402–415. [CrossRef]

62. Habtewold, T.D.; Islam, A.; Liemburg, E.J.; Bruggeman, R.; Alizadeh, B.Z.; Bartels-Velthuis, A.A.; van Beveren, N.J.; Cahn, W.; de Haan, L.; Delespaul, P.; et al. Polygenic risk score for schizophrenia was not associated with glycemic level (HbA1c) in patients with non-affective psychosis: Genetic Risk and Outcome of Psychosis (GROUP) cohort study. *J. Psychosom. Res.* **2020**, *132*, 109968. [CrossRef]

63. McElreath, R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2020.

64. Li, Y.; Lord-Bessen, J.; Shiyko, M.; Loeb, R. Bayesian Latent Class Analysis Tutorial. *Multivar. Behav. Res.* **2018**, *53*, 430–451. [CrossRef]

65. Asparouhov, T.; Muthén, B. Using Bayesian priors for more flexible latent class analysis. In Proceedings of the 2011 Joint Statistical Meeting, Miami Beach, FL, USA, 4 August 2011; American Statistical Association: Alexandria, VA, USA, 2011.

66. Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput. Sci.* **2021**, *2*, 160. [CrossRef] [PubMed]