Rethinking the Course of Psychotic Disorders: Modelling Long-Term Symptom Trajectories

Craig Morgan PhD[1,2], Paola Dazzan MRCPsych[2,10], Julia Lappin MRCPsych[11], Margaret Heslin PhD[4], Kim Donoghue PhD[5], Paul Fearon MRCPsych[6], Peter B Jones FRCPsych[7], Robin M Murray FRS[2,3], Gillian A Doody FRCPsych[8], and Ulrich Reininghaus PhD[1,9]

Correspondence to: Professor Craig Morgan, ESRC Centre for Society and Mental Health, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK. Tel: 020 7848 0351. E-mail: craig.morgan@kcl.ac.uk

[1]ESRC Centre for Society and Mental Health, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK; [2]National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; [3]Psychosis Studies Department, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK; [4]King's Health Economics, Health Service and Population Research Department, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK; [5]Addictions Department, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK; [6]Department of Psychiatry, Trinity College, Dublin, Ireland; [7]Department of Psychiatry, University of Cambridge, Cambridge, UK; [8]Division of Psychiatry and Applied Psychology, University of Nottingham, Nottingham, UK; [9]Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany; [10]Department of Psychological Medicine, Institute of Psychiatry, Psychology, and Neuroscience, King's College, London, UK; [11]Faculty of Medicine, University of New South Wales, Sydney, Australia.

1

Running Title   Rethinking the Course of Psychotic Disorders

Abstract Word Count   242

Text Body Word Count   4500

Tables and Figures   5

Abstract

Background

The clinical course of psychotic disorders is highly variable. Typically, researchers have captured different course types using broad pre-defined categories. However, whether these adequately capture symptom trajectories of psychotic disorders has not been fully assessed. Using data from AESOP-10, we sought to identify classes of individuals with specific symptom trajectories over a 10-year follow-up using a data driven approach.

Method

AESOP-10 is a follow-up, at 10 years, of 532 incident cases with a first episode of psychosis initially identified in south-east London and Nottingham, UK. Using extensive information on fluctuations in presence of psychotic symptoms, we fitted growth mixture models to identify latent trajectory classes that accounted for heterogeneity in patterns of change in psychotic symptoms over time.

Results

We had sufficient data on psychotic symptoms during the follow-up on 326 incident patients. A four-class quadratic growth mixture model identified four trajectories of psychotic symptoms: (1) remitting-improving (58.5%); (2) late decline (5.6%); and (3) late improvement (5.4%); (4) persistent (30.6%). A persistent trajectory, compared with remitting-improving, was associated with gender (more men), black Caribbean ethnicity, low baseline education and high disadvantage, low premorbid IQ, a baseline diagnosis of non-affective psychosis, and long DUP. Numbers were small, but there were indications that those with a late decline trajectory more closely resembled those with a persistent trajectory.

Conclusion

Our current approach to categorising course of psychotic disorders may misclassify patients. This may confound efforts to elucidate predictors of long-term course and related biomarkers.

Keywords Psychotic Disorder, Course and Outcome, Epidemiology, Growth Mixture Models

Introduction

Our knowledge of the nature and predictors of the long-term course and outcomes of psychotic disorders is surprisingly limited. In a previous attempt to synthesise the evidence, we identified only 13 studies, published up to 2014, that had followed individuals with a first episode of psychosis and reported on course and outcome over a period of 8 years or more (Morgan et al., 2014). We are aware of two further studies since then (Kotov et al., 2017; Secher et al., 2015). This body of research is methodologically varied, which makes direct comparisons difficult and limits any general conclusions we can make about the long-term trajectories of psychosis following a first episode and associated factors.

Further, our characterisation of the symptomatic course of psychotic disorders is mostly crude and static. Following Ciompi (Ciompi, 1980), researchers have tended to characterise symptomatic course by assigning individuals to broad groups based on lengths of episodes and remissions (Andreasen et al., 2005; Harrison et al., 2001; Hopper, Harrison, Janca, & Sartorius, 2007; Moller et al., 2010; White et al., 2009). For example, in one common characterisation, an episodic course is defined as one in which no episode of psychosis lasted 6 months or more; a continuous course as one in which no remission lasted 6 months or more; and neither, by default, as one in which there was an episode of 6 months or more and a remission of 6 months or more (e.g., (Hopper et al., 2007)). The validity of this approach has not, as far as we are aware, been assessed and, by using a pre-defined cut point of 6 months to define episodes and remissions, it may misclassify individuals with similar overall trajectories and consequently confound efforts to identify predictors of course and outcome.

However, developing more fine-grained approaches to investigating the long-term course of psychotic disorders presents several methodological challenges. Even categorising course as above requires reconstructing periods of relapse and remission over many years. This said, use of multiple sources and informants, life course interview techniques, and consensus ratings can improve the validity of retrospective reconstructions of periods of symptoms (Bebbington et al., 2006). However, analysing detailed data on fluctuations in symptoms over many years presents further challenges and this probably explains why these fluctuations are usually distilled into a small number of static categories. There are statistical methods that offer an alternative, by enabling a data driven approach to identifying individuals whose symptoms follow similar trajectories over time (Austin et al., 2015; B. Muthen, 2004; B. Muthen & Muthen, 2000). For example, latent growth mixture models can identify

trajectories (i.e. categorical latent variables) that account for heterogeneity in fluctuations of psychotic symptoms over time (Grimm, Ram, & Estabrook, 2017). Such an approach may produce more valid groupings and enable us to better elucidate predictors of course and outcome.

Using extensive data from AESOP-10, collated from multiple sources and meticulously rated by consensus, we sought to examine fluctuations in symptoms over a 10-year follow-up period. Specifically, we sought to identify classes of individuals with specific symptom trajectories over a 10-year follow-up and to, then, compare trajectories with usual categories of course and examine associations between trajectories and baseline demographic, social, and clinical characteristics.

Method

AESOP-10 is a follow-up at approximately 10 years of a cohort of 532 incident cases with a first episode of non-affective or affective psychosis initially identified in South East London and Nottingham, UK. There were marked differences in incidence between the two sites (i.e., adj. incidence rate 49.4 (95% CI 43.6-55.3) per 100,000 person-years in London and 23.9 (95% CI 20.6-27.2) per 100,000 in Nottingham. See Kirkbride et al., 2006).

Ethics

Ethical approval was provided by the Joint South London and Maudsley and Institute of Psychiatry NHS Research Ethics Committee (Ref: 321/02) and the North Nottinghamshire Local Research Ethics Committee (Ref: 04/Q2402/35). The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Follow-up Procedures

At baseline, detailed re-contact information was collected for all patients. At approximately 10-years, we sought to re-contact and re-interview each patient. All deaths and emigrations up to 12 December 2012 were identified through the Office for National Statistics (ONS) for England and Wales and the General Register Office (GRO) for Scotland (full details on mortality in the AESOP-10 cohort can be found in Reininghaus et al (2015)).

Data (1) Baseline

At baseline, all individuals with a first-episode psychotic disorder who presented to secondary mental health services in our catchment areas over two years were identified from 1997-1999 (i.e., before Early Intervention Services were introduced in the UK). This predates the introduction of early intervention services in the UK and all consequently received standard inpatient and community care. Data were collected on clinical presentation, sociodemographic characteristics, and a range of neurodevelopmental and social risk factors. We estimated premorbid Intelligence Quotient (IQ) using the National Adult Reading Test (NART; Nelson, 1991). To capture exposure to multiple disadvantages and isolation, we constructed an index by counting the presence of the following: unemployment, living alone, living in rented housing, and being single.

Data (2) Follow-up

At follow-up, information was collated across three domains (clinical, social, service use) using an extended version of the WHO Life Chart (Susser et al., 2000) and the Global Assessment of Function Disability (GAF-D) Scale (Endicott, Spitzer, Fleiss, & Cohen, 1976). We extended the WHO Life Chart to include a timeline to record the presence and absence of psychotic symptoms month by month during the follow up period. Information on psychotic symptoms was collected from interviews with patients and, with consent, from informants (e.g., clinicians, relatives) and case records (Morgan et al., 2014). Interviews used life course techniques, including anchoring around key events (e.g., birthdays, hospital admissions, etc.), to maximise recall. Using all information from follow-up interviews with patients and treating clinicians (informants) and from clinical records, researchers methodically reconstructed patient histories over the follow-up period. Ratings of presence or absence of psychotic symptoms were based on Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (WHO, 1992) criteria for rating of clinically important symptoms and were only made on the basis of clear information indicating presence or absence (i.e., clear description of occurrence of symptoms during a specified period at a sufficient level of intensity to be rated present according to SCAN criteria); otherwise, data were coded as missing. All final ratings of presence, absence, or missing (i.e., insufficient information to make a rating) were made by consensus at weekly meetings involving members of the research team and a senior psychiatrist. In addition, information on substance abuse and dependence and markers of social disadvantage and isolation across a number of domains during and at follow-up was collected using the Life Chart. The GAF-D Scale, which provides information on overall social function, was completed for the one-month pre-follow-up.

Analysis

Growth mixture modelling was used to identify latent trajectory classes (i.e. categorical latent variables) that account for heterogeneity in patterns of change in psychotic symptoms over time (Grimm et al., 2017; B. Muthen, 2004; B. Muthen & Aspharouhov, 2008). We fitted a series of models to determine the classification that best represents temporal patterns of change in psychotic symptoms. First, we fitted growth mixture models (GMM) without random intercept and slope (B. Muthen, 2004; Nagin, 2005). We then added quadratic and cubic terms of time to account for potential non-linear growth in the data. We also examined whether adding random intercepts and slopes that account for heterogeneity in symptom change within latent trajectory classes further improved model fit. Data were assumed to be missing at random (MAR), which allowed for computing unbiased estimates based on all available data using maximum likelihood estimation with robust standard errors (i.e., the MLR estimation method in MPlus). To assess the assumption of MAR, we examined the association between socio-demographic and clinical characteristics at baseline and missingness in timeline data over 10 years of follow-up fitting random-intercept logistic regression models. GMM fit was assessed using the Bayesian Information Criterion (BIC) (Schwarz, 1978), with lower values and differences equal to or above 10 indicating better model fit than for the comparison model (Raftery, 1995). We computed the entropy for models with more than one class to examine the quality of classification accuracy. Entropy values closer to 1 indicate a better distinction between latent classes (range 0-1). We also computed average latent class probabilities for most likely latent class membership as another indicator of goodness of model fit or uncertainty in assigning individuals to latent classes given the observed data (Geiser, 2013). GMM were further compared using the Lo-Mendell-Rubin likelihood ratio test (Lo, Mendell, & Rubin, 2001) and a bootstrapped likelihood ratio test with 500 draws (Nylund, Aspharouhov, & Muthen, 2007) to examine whether models with additional classes improved model fit (as indicated by $P < 0.05$), compared with a model with one less class. In determining the number of classes in the final model, we also considered interpretability and class sizes (Uher et al., 2010). Specifically, we only included additional latent classes if they represented fundamentally different trajectories rather than minor variations of those included in the previous model (B. Muthen, 2004). Further, we aimed for a final model without latent class trajectories including only a very small proportion of subjects (< 5%). Trajectories including equal to or more than 5% of subjects were still considered as clinically important, given that, identifying and targeting predictors of trajectories of symptom change even in relatively small

7

subpopulations can lead to potentially substantial public health benefits. Analyses were conducted using STATA Version 12 and MPlus Version 7.1. (L. K. Muthen & Muthen, 2013).

Results

Sample

At baseline, we identified 532 incident cases, of whom 390 consented to complete assessments. At follow up, of the 532 incident cases, 37 had died, 29 had emigrated, and 8 were excluded. Of the remaining 458, we were able to complete a timeline of psychotic symptoms throughout the follow up for 326 (71% of 458), with a mean length of follow-up of 10.6 years (SD 1.1). Characteristics of the included sample (n, 326) vs. those not included (n, 206) are shown in Supplementary Table 1.

Missing Data

There was no evidence of an association between missingness and age, gender, ethnicity, level of education, social disadvantage, pre-morbid IQ, diagnosis, DUP, and mode of onset, and only weak evidence of an association between duration of untreated psychosis and missingness in timeline data over the 10-year follow-up (see Supplementary Table 2). This suggests the assumption that symptom data are missing at random (MAR) is acceptable.

Trajectories of Symptom Change in Psychotic Symptoms

Overall, GMM with random intercept and slope (Supplementary Table 4) showed a better model fit than GMM without random intercepts and slopes (Supplementary Table 3). GMM with linear effects provided a good model fit (Supplementary Table 4, Models 1.1.1 to 1.1.6). However, estimated means of linear time effects in these models differed markedly from observed data and model fit was further improved when adding quadratic effects of time to account for non-linear growth in the observed data, with high quality of classification of subjects into latent trajectory classes (entropy > 0.90). A lower BIC value was found for GMM with quadratic effects (Supplementary Table 4, Models 1.2.1-1.2.6) compared with GMM with linear effects (Supplementary Table 4, Models 1.1.1-1.1.6), whereas GMM with cubic effects yielded a non-positive definite residual covariance matrix and first-order derivative product matrix even after increasing starting values and, hence, pointed to model non-identification.

The BIC and Lo-Mendell-Rubin test indicated a markedly improved model fit for the quadratic GMM with two latent trajectory classes (Supplementary Table 4; Supplementary Figures 1 & 2). However, there was still marked departure of observed temporal patterns of changes in psychotic symptoms from estimated means in this quadratic, 2-class GMM (Supplementary Table 5). In line with this, the BIC and bootstrapped likelihood ratio test indicated that model fit continued to improve for the 3-, 4-, 5-, and 6-class quadratic GMM (Supplementary Table 4). However, there was a lower entropy for the 5- and 6-class quadratic GMM compared with the 4-class quadratic GMM. Given also that the 5- and 6-class quadratic GMM included trajectories with less than 5% of subjects, the 4-class quadratic GMM was selected as the final, best-fitting model (see Figures 1 & 2; Supplementary Tables 4 & 6), in which average latent class probabilities for most likely class membership was very high (i.e., >0.98; see Supplementary Table 6). The classes can be characterised as remitting-improving (Class 1, 58.5%), late decline (Class 2, 5.6%), late improvement (Class 3, 5.4%), and persistent (Class 4, 30.6%) trajectories.

[Insert Figures 1, 2]

Latent Trajectories and Other Course and Outcome Variables

When compared with the usual classification of course types (episodic, continuous, and neither), the two largest classes captured all those in the episodic (n 94 of 94; 100%) and almost all those in the continuous (n 72 of 78; 92%) categories. A majority of those in the neither category were classified in the remitting class (i.e., n 90 of 145; 62%), with the remainder spread across the other three classes (i.e., n 25, 17% persistent; n 14, 10% late decline; n 16, 11% late improvement) (Table 1). With regard to symptom recovery (i.e., been symptom free for the preceding two years), almost three quarters (73%; 124 of 179) of those in the remitting class had recovered symptomatically at follow up. By definition, none of those in the continuous class or the late decline class had recovered. Among the small class of individuals whose symptoms improved late, more than half (56%; 9 of 17) had recovered at follow up.

[Insert Table 1]

Further, there were strong associations between latent symptom trajectories and social outcomes (Table 2). For example, the mean GAF disability score at follow up was almost 20 points higher among those in the remitting class compared with those in the persistent class (mean GAF-D: remitting 63.7 [sd 17.9] vs. persistent 45.9 [sd 12.2]; F 30.0, df 3, p < 0.001). The mean scores for the late decline (41.2; sd 17.4) and improvement (49.8; sd 12.4) classes were similar to that for the persistent group (45.9; sd 12.2). On broad markers of social exclusion, over 60% of those in the remitting class were not in a long-term relationship at follow up and over 60% were unemployed. These proportions were, not surprisingly, lower still among those in the persistent class, particularly for those employed (less than 10%). Even so, around 20% who were more or less continuously symptomatic were in a relationship, which underscores the importance of considering symptomatic and social outcomes separately.

[Insert Table 2]

Demographic, Social, and Clinical Characteristics

There were several strong associations between baseline demographic, social, and clinical variables, which closely mirror previous findings (Table 3 and Supplementary Table 7). For example, compared with the remitting class, in the persistent class there were more men (Adj. OR 1.89, 95% CI 1.13-3.17), more from London (Adj. OR 1.58, 95%CI 0.84-2.96), more of black Caribbean ethnicity (vs. white British, Adj. OR 1.96, 95% CI 1.01-3.80), more with lower levels of education (e.g., school vs. university, Adj. OR 3.86, 95% CI 1.20-12.39) and with higher levels of social disadvantage and isolation (e.g., 4 markers vs 0 or 1, Adj. OR 2.72, 95% CI 1.08-6.88), more in the lowest quantile of premorbid IQ (e.g., lowest vs highest quartile, Adj. OR 3.35 (0.82-13.50), fewer with a diagnosis of a affective psychotic disorder (Adj. OR 0.17, 95% CI 0.08-0.38), and DUP was substantially longer (Adj. OR for each additional month DUP 1.05, 95% CI 1.01-1.09). Not surprisingly, in the persistent class there were more who were treatment resistant from the outset (n 36 of 46 (78%); Supplementary Table 8). There was no strong evidence of differences in substance abuse or dependence among those in the persistent vs. remitting class.

[Insert Table 3]

In the other two classes - late improvement and decline - the numbers mean it is difficult to discern any clear patterns. Nonetheless, two observations are worth noting. First, those in the late improvement class (i.e., those who gradually improve from around 3 years) had a much shorter DUP, on average, and were more likely to meet criteria for substance abuse or dependence at some point than those in the persistent class. Second, there were some indications that those with a late decline trajectory more closely resembled those with a persistent trajectory (i.e., more of black Caribbean ethnicity, more at higher levels of social disadvantage, fewer with a diagnosis of affective psychosis, and longer DUP compared with remitting) than did those with a late improvement trajectory (i.e., no evidence of demographic and diagnostic differences compared with remitting). Still, given the small numbers, this reading of the data is somewhat speculative and needs to be considered with caution, as the basis for hypotheses to be tested in larger samples.

Discussion

We sought to use a data driven approach to identify classes of change in psychotic symptoms over time. Symptom course was highly variable. This noted, we identified four latent trajectory classes. Two were common: i.e., a trajectory characterised by remitting periods of symptoms, which became shorter and less frequent over time, leading to recovery for many (~ 60%), and a trajectory characterised by persistent or long periods of symptoms over the whole follow up (~ 30%), with many meeting treatment resistant criteria. For most, the pattern of course changes – and improves – over time, a finding in line with the only other attempt we are aware of to use latent class models to capture course types (albeit, this study did not model growth – i.e. change – over time).(Austin et al., 2015) Overall, this suggests slightly more positive symptom outcomes than some studies (e.g., International Study of Schizophrenia, ~ 50% recovered at 15 year follow-up (Harrison et al, 2002) and substantially more positive outcomes than in some others (e.g., Suffolk County Study, ~ 25% with non-continuous illness at 20 year follow-up (Kotov et al., 2017)). That patterns of symptoms do change over time is further highlighted by the other two classes: late decline (~ 5%) and late improvement (~ 5%). Further, our analyses identified several baseline characteristics that were associated with a persistent course, i.e., men, black Caribbean ethnicity, urban context (i.e., London), social disadvantage, low premorbid IQ, long DUP, and non-affective psychoses.

Methodological Considerations

11

Our findings need to be considered in light of a several methodological limitations. In attempting to retrospectively reconstruct fluctuations in symptoms over a 10-year period, measurement error and missing data are inevitable. We used a number of approaches to minimise measurement error and maximise the reliability of data, including use of life course interview techniques, multiple sources and informants, and consensus ratings, and only made ratings based on positive evidence of presence or absence of symptoms. However, we were reliant on recall and use of clinical records, which varied in detail and periods covered. This further increased the amount of missing data. This is relevant at two levels. First, a proportion of patients for whom we did not have sufficient data over the follow up period could not be included in our analyses. Those with sufficient data and included in analyses were more likely to be from London (vs. Nottingham) and less likely to be white British. This may have introduced bias. It is possible, for example, that those who were not included may have had better outcomes, on the basis that they had fewer contacts with mental health services, which would bias our findings toward poor outcomes. Conversely, it is possible that those with negative symptoms had fewer contacts with services, which would bias our findings toward better outcomes. Second, data on presence or absence of symptoms did not include broader measures of outcome such as social functioning, which may have yielded different trajectories as, for example, those identified by Velthorst et al. (2017). Further, WHO Life Chart data were collected for the 10-year follow up period (i.e., first presentation to follow-up) and not for the period before baseline (i.e., symptom onset to first presentation). Consequently, we cannot rule out that some individuals may have been misclassified into the late decline or persistent trajectories because symptom data were not available for the period before first presentation. Third, of the patients included in the sample, not all had complete data for the entire follow up period. As indicated already, with the possible exception of DUP, there was no evidence of associations between missingness in the timeline data and sociodemographic and clinical characteristics. From this, we were able to generate unbiased estimates based on all available data, on the assumption that data were missing at random. Still, these issues provide important caveats to our data and what can be inferred.

Rethinking Course: Implications

These limitations noted, our analyses provide a novel approach to studying the course of disorder and offer potentially important insights.

To begin with, our analyses challenge our way of conceptualising course of psychotic disorders in at least three ways. First, they suggest use of any pre-specified cut points for remission and relapse to distinguish course types is overly crude and leads to misclassification that may confound efforts to elucidate predictors of course. Second, our findings suggest the use of a middle category of neither, between episodic and continuous, is flawed and may confound findings. This is supported by our previous analyses of MRI data on a sub-sample of AESOP cases and controls using a support vector machine whole-brain classification approach to predict future illness course at the individual level (Mourao-Miranda et al., 2012). Broadly, we found that those in the neither episodic nor continuous group could mostly be categorised according to a discriminating pattern of episodic vs continuous. Third, the use of static categories to capture course implies that individuals follow a single trajectory that is unchanging over time. Our data suggest that this is true, at most, for a minority who experience (near) persistent symptoms. For others, the course fluctuates. A majority tend to experience shorter and shorter episodes, a feature that is missed in the simple category of episodic. A small, but noteworthy, proportion experience marked changes in the course of symptoms over time, 3 to 5 years after a first episode. These groups are important because they suggest improvement, after an initially poor course, is possible and because they point to a group of individuals who later experience long periods of psychosis, after initial short periods. These findings are likely to be intuitively valid to clinicians who regularly see patients with symptoms that wax and wane over time, and who may disengage from services at times when symptoms are absent. This highlights the need for clinical teams to regularly monitor for psychotic symptoms and the need for fluid access to, and exit from, mental health services, with re-entry to specialist services available at times when symptoms recur.

As noted, we found several broad demographic, social, and clinical variables were associated with course and outcome over the long-term. However, our ability to predict long-term prognosis – particularly at an individual level in clinical practice – is still limited (Suvisaari et al., 2018). Diagnosis provides perhaps the best clinical indicator of prognosis – as we would expect – but there remains considerable heterogeneity in outcomes within diagnostic groups. Our findings hint at several reasons for this. For example, potential issues with measurement error and missing data notwithstanding, our findings highlight the marked variability in fluctuations in symptoms over time. We can discern broad trends, as our identified classes illustrate, and associated factors, but it remains that what lies behind these is considerable heterogeneity in patterns of remission and relapse over time. This is intrinsically hard to predict. Further, it seems that baseline predictors of outcomes may weaken over time.(P

Dazzan et al., In press) There is a considerable literature on prognostic markers for short to medium term outcomes, including substance use (especially cannabis) (Bozzatello, Bellino, & Rocca, 2019; Schoeler et al., 2016), neurocognition (Kravariti et al., 2019; Lepage, Bodnar, & Bowie, 2014), negative symptoms (Suvisaari et al., 2018), and various biomarkers (e.g., brain structure) (P. Dazzan et al., 2015; McGuire & Dazzan, 2017). However, findings are often inconsistent and it is unclear to what extent these predict longer-term outcomes. Indeed, putting our findings in this paper together with findings from other analyses of AESOP-10 data, there is a relatively small number of baseline demographic, social, and clinical characteristics that are associated with long-term course and outcome and the strongest predictor of later outcomes is early outcomes (P Dazzan et al., In press; Demjaha et al., 2017; Lappin et al., 2014). This hints at a dynamic process in which initial drivers of onset and early course interact with early outcomes to shape subsequent exposure to intervening factors that – over time – are more important in shaping course (e.g., treatment response, substance use, social circumstances, etc.), creating dynamic feedback loops that lead – for most – to idiosyncratic patterns of symptom remission and relapse. At present, we do not have the studies, with sufficient depth of data, to enable more dynamic models to be tested.

Of particular interest from our analyses are those whose trajectories change notably late in the course of illness. Those who experience a late decline appear similar at baseline demographically and clinically to those who experience near persistent symptoms (e.g., longer DUP). The small number in this class urges caution, but if validated in subsequent studies this has potentially important clinical implications and suggests those with, for example, a long DUP may need to be more carefully assessed and treated over time, even if the initial trajectory is positive. Further, those who improve and recovery late in the course of their illness may shed light on what can be done to promote recovery when the initial course is poor. There are few clues to this in our data, because the number in this group was small, but it is an important potential avenue for future research. Particular consideration, for example, should be given to the potential impacts on improving later outcomes of clozapine, reduced substance use, and greater social inclusion.

Conclusion

Symptomatic course, following a first-episode psychosis, is heterogenous and intrinsically difficult to predict, especially over the long-term. This poses considerable challenges for both research and clinical practice. With

the advent of electronic clinical records in many services around the world and the development of statistical methods that enable more sophisticated modelling of change over time, illustrated by analyses presented in this paper, there are opportunities for a new wave of research on course and outcome in psychotic disorders. Of particular importance, there is a substantial sub-group – as ours and other studies highlight – that experiences near persistent symptoms throughout, often despite treatment with antipsychotic medication. The associated suffering, needs, and costs are hard to over-state. This underscores the importance of investment both in research and in the continued development of interventions and services to better understand, predict, and support those whose symptoms persist over the long-term despite current treatment options.

Conflict of interest statement

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

References

Andreasen, N. C., Carpenter, W. T., Jr., Kane, J. M., Lasser, R. A., Marder, S. R., & Weinberger, D. R. (2005). Remission in schizophrenia: proposed criteria and rationale for consensus. *American Journal of Psychiatry, 162*(3), 441-449.

Austin, S. F., Mors, O., Budtz-Jorgensen, E., Secher, R. G., Hjorthoj, C. R., Bertelsen, M., . . . Nordentoft, M. (2015). Long-term trajectories of positive and negative symptoms in first episode psychosis: A 10year follow-up study in the OPUS cohort. *Schizophrenia Research, 168*(1-2), 84-91.

Bebbington, P. E., Craig, T., Garety, P., Fowler, D., Dunn, G., Colbert, S., . . . Kuipers, E. (2006). Remission and relapse in psychosis: operational definitions based on case-note data. *Psychological Medicine, 36*(11), 1551-1562.

Bozzatello, P., Bellino, S., & Rocca, P. (2019). Predictive Factors of Treatment Resistance in First Episode of Psychosis: A Systematic Review. *Frontiers in Psychiatry, 10*, 67.

Ciompi, L. (1980). The natural history of schizophrenia in the long term. *British Journal of Psychiatry, 136*, 413-420.

Dazzan, P., Arango, C., Fleischacker, W., Galderisi, S., Glenthoj, B., Leucht, S., . . . McGuire, P. (2015). Magnetic resonance imaging and the prediction of outcome in first-episode schizophrenia: a review of current evidence and directions for future research. *Schizophrenia Bulletinl, 41*(3), 574-583.

Dazzan, P., Lappin, J., Heslin, M., Donoghue, K., Lomas, B., Reininghaus, U., . . . Morgan, C. (In press). Symptom remission at 12-weerks strongly predicts long-term recovery from the first episode of psychosis. *Psychological Medicine*.

Demjaha, A., Lappin, J. M., Stahl, D., Patel, M. X., MacCabe, J. H., Howes, O. D., . . . Murray, R. M. (2017). Antipsychotic treatment resistance in first-episode psychosis: prevalence, subtypes and predictors. *Psychological Medicine, 47*(11), 1981-1989.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*(6), 766-771.

Geiser, C. (2013). Data analysis with MPlus. New York: The Guilford Press.

Grimm, K. J., Ram, N., & Estabrook, R. (2017). *Growth modeling : structural equation and multilevel modeling approaches*. New York, NY: Guilford Press.

Harrison, G., Hopper, K., Craig, T., Laska, E., Siegel, C., Wanderling, J., . . . Wiersma, D. (2001). Recovery from psychotic illness: a 15- and 25-year international follow-up study. *British Journal of Psychiatry, 178*, 506-517.

Hopper, K., Harrison, G., Janca, A., & Sartorius, N. (Eds.). (2007). *Recovery from schizophrenia: An international perspective*. Oxford: Oxford University Press.

Kirkbride, J. B., Fearon, P., Morgan, C., Dazzan, P., Morgan, K., Tarrant, J., . . . Jones, P. B. (2006). Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: findings from the 3-center AeSOP study. *Archives of General Psychiatry, 63*(3), 250-258.

Kotov, R., Fochtmann, L., Li, K., Tanenberg-Karant, M., Constantino, E. A., Rubinstein, J., . . . Bromet, E. J. (2017). Declining Clinical Course of Psychotic Disorders Over the Two Decades Following First Hospitalization: Evidence From the Suffolk County Mental Health Project. *American Journal of Psychiatry, 174*(11), 1064-1074.

Kravariti, E., Demjaha, A., Zanelli, J., Ibrahim, F., Wise, C., MacCabe, J. H., . . . Murray, R. M. (2019). Neuropsychological function at first episode in treatment-resistant psychosis: findings from the AESOP-10 study. *Psychological Medicine, 49*(12), 2100-2110.

Lappin, J. M., Morgan, C., Chalavi, S., Morgan, K. D., Reinders, A. A., Fearon, P., . . . Dazzan, P. (2014). Bilateral hippocampal increase following first-episode psychosis is associated with good clinical, functional and cognitive outcomes. *Psycholological Medicine, 44*(6), 1279-1291.

Lepage, M., Bodnar, M., & Bowie, C. R. (2014). Neurocognition: clinical and functional outcomes in schizophrenia. *Canadian Journal of Psychiatry, 59*(1), 5-12.

Lo, Y., Mendell, N., & Rubin, D. (2001). Tsting the number of components in a normal mixture. *Biometrika, 88*, 767-778.

McGuire, P., & Dazzan, P. (2017). Does neuroimaging have a role in predicting outcomes in psychosis? *World Psychiatry, 16*(2), 209-210.

Moller, H. J., Jager, M., Riedel, M., Obermeier, M., Strauss, A., & Bottlender, R. (2010). The Munich 15-year follow-up study (MUFUSSAD) on first-hospitalized patients with schizophrenic or affective disorders: comparison of psychopathological and psychosocial course and outcome and prediction of chronicity. *European Archives of Psychiatry and Clinical Neuroscience, 260*(5), 367-384.

Morgan, C., Lappin, J., Heslin, M., Donoghue, K., Lomas, B., Reininghaus, U., . . . Dazzan, P. (2014). Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. *Psychological Medicine, 44*(13), 2713-2726.

Mourao-Miranda, J., Reinders, A. A. T. S., Rocha-Rego, V., Lappin, J., Rondina, J., Morgan, C., . . . Dazzan, P. (2012). Individualized prediction of illness course at the first psychotic episode: a support vector machine MRI study. *Psychological Medicine, 42*(5), 1037-1047.

Muthen, B. (2004). Latent variable analysis. In D. Kaplan (Ed.), *SAGE handbook of quantitative methodology for the social sciences* (pp. 345-368). London: SAGE Publications.

Muthen, B., & Aspharouhov, T. (2008). Growth mixture modelling: analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal Data Analysis* (pp. 143-165). USA: CRC Press.

Muthen, B., & Muthen, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891.

Muthen, L. K., & Muthen, B. (2013). Mplus Version 7.1. Los Angeles, CA.

Nagin, D. (2005). *Group-based modeling of development*. Cambridge, Mass.: Harvard University Press.

Nelson, H.W. (1991). *National Adult Reading Test (Part II) Test Manual.* Windsor, Canada: NFER-NELSON.

Nylund, K. L., Aspharouhov, T., & Muthen, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Structural Equation Modeling, 14*, 535-569.

Raftery, A. (1995). Bayesian model selection in social research. In P. Marsden (Ed.), *Sociological Methodology* (pp. 111-163). UK: Blackwell.

Reininghaus, U., Dutta, R., Dazzan, P., Doody, G.A., Fearon, P., Lappin, J., Heslin, M., Onyejiaka, A., Donoghue, K., Lomas, B., Kirkbride, J.B., Murray, R.M., Croudace, T., Morgan, C., Jones, P.B. (2015). Mortality in schizophrenia and other psychoses: a 10-year follow-up of the ÆSOP first episode cohort. *Schizophrenia Bulletin, 41*, 664-673.

Schoeler, T., Monk, A., Sami, M. B., Klamerus, E., Foglia, E., Brown, R., . . . Bhattacharyya, S. (2016). Continued versus discontinued cannabis use in patients with psychosis: a systematic review and meta-analysis. *Lancet Psychiatry, 3*(3), 215-225.

Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461-464.

Secher, R. G., Hjorthoj, C. R., Austin, S. F., Thorup, A., Jeppesen, P., Mors, O., & Nordentoft, M. (2015). Ten-year follow-up of the OPUS specialized early intervention trial for patients with a first episode of psychosis. *Schizophrenia Bulletin, 41*(3), 617-626.

Susser, E., Finnerty, M., Mojtabai, R., Yale, S., Conover, S., Goetz, R., & Amador, X. (2000). Reliability of the life chart schedule for assessment of the long-term course of schizophrenia. *Schizophrenia Research, 42*(1), 67-77.

Suvisaari, J., Mantere, O., Keinanen, J., Mantyla, T., Rikandi, E., Lindgren, M., . . . Raij, T. T. (2018). Is It Possible to Predict the Future in First-Episode Psychosis? *Frontiers in Psychiatry, 9*, 580.

Uher, R., Muthen, B., Souery, D., Mors, O., Jaracz, J., Placentino, A., . . . McGuffin, P. (2010). Trajectories of change in depression severity during treatment with antidepressants. *Psychological Medicine, 40*(8), 1367-1377.

Velthorst, E., Fett, A. J., Reichenberg, A., Perlman, G., van Os, J., Bromet, E. J., & Kotov, R. (2017). The 20-Year Longitudinal Trajectories of Social Functioning in Individuals With Psychotic Disorders. *American Journal of Psychiatry*, *174*(11), 1075–1085.

White, C., Stirling, J., Hopkins, R., Morris, J., Montague, L., Tantam, D., & Lewis, S. (2009). Predictors of 10-year outcome of first-episode psychosis. *Psychological Medicine, 39*(9), 1447-1456.

WHO. (1992). *Schedules for Clinical Assessment in Neuropsychiatry*. Geneva: World Health Organisation.