The clinical significance of duration of untreated psychosis: an umbrella review and random-effects meta-analysis

Oliver D. Howes[1-4], Thomas Whitehurst[2,3], Ekaterina Shatalina[2,3], Leigh Townsend[2,3], Ellis Chika Onwordi[1-4], Tsz Lun Allenis Mak[5], Atheeshaan Arumuham[1-4], Oisín O'Brien[1], Maria Lobo[1], Luke Vano[1], Uzma Zahid[1], Emma Butler[1,4], Martin Osugo[1-3]

[1]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]MRC London Institute of Medical Sciences, Hammersmith Hospital, London, UK; [3]Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London, UK; [4]South London and Maudsley Foundation NHS Trust, Maudsley Hospital, London, UK; [5]St. George's, University of London, London, UK

The idea that a longer duration of untreated psychosis (DUP) leads to poorer outcomes has contributed to extensive changes in mental health services worldwide and has attracted considerable research interest over the past 30 years. However, the strength of the evidence underlying this notion is unclear. To address this issue, we conducted an umbrella review of available meta-analyses and performed a random-effects meta-analysis of primary studies. MEDLINE, Web of Science, PsycINFO and EMBASE were searched from inception to September 3, 2020 to identify relevant meta-analyses of studies including patients with schizophrenia spectrum disorders, first-episode psychosis, or affective and non-affective psychosis. Thirteen meta-analyses were included, corresponding to 129 individual studies with a total sample size of 25,657 patients. We detected potential violations of statistical assumptions in some of these meta-analyses. We therefore conducted a new random-effects meta-analysis of primary studies. The association between DUP and each outcome was graded according to a standardized classification into convincing, highly suggestive, suggestive, weak, or non-significant. At first presentation, there was suggestive evidence for a relationship between longer DUP and more severe negative symptoms (beta=−0.07, p=3.6×10^{−5}) and higher chance of previous self-harm (odds ratio, OR=1.89, p=1.1×10^{−5}). At follow-up, there was highly suggestive evidence for a relationship between longer DUP and more severe positive symptoms (beta=−0.16, p=4.5×10^{−8}), more severe negative symptoms (beta=−0.11, p=3.5×10^{−10}) and lower chance of remission (OR=2.16, p=3.0×10^{−10}), and suggestive evidence for a relationship between longer DUP and poorer overall functioning (beta=−0.11, p=2.2×10^{−6}) and more severe global psychopathology (beta=−0.16, p=4.7×10^{−6}). Results were unchanged when analysis was restricted to prospective studies. These effect sizes are clinically meaningful, with a DUP of four weeks predicting >20% more severe symptoms at follow-up relative to a DUP of one week. We conclude that DUP is an important prognostic factor at first presentation and predicts clinically relevant outcomes over the course of illness. We discuss conceptual issues in DUP research and methodological limitations of current evidence, and provide recommendations for future research.

Key words: Duration of untreated psychosis, outcomes, negative symptoms, positive symptoms, schizophrenia, remission, functioning, global psychopathology, recommendations for research

Psychotic disorders such as schizophrenia are often marked by persistent symptoms, reduced quality of life, and long-term disability[1]. There have been few advances in drug treatment in the past 30 years, with a concomitant growth of interest in modifiable factors which may determine outcomes[2]. The 1986 Northwick Park study highlighted that some patients with psychosis experienced considerable delays before starting treatment, and that this delay was associated with poorer outcomes once treatment was initiated[3]. This was subsequently conceptualized as the duration of untreated psychosis (DUP), which is generally considered to be the period from onset of psychotic symptoms to the initiation of treatment[4].

It was later proposed that psychosis has a persistent neurotoxic effect which cannot be fully reversed even once treatment is initiated[5]. The critical window hypothesis extended this concept to suggest that deterioration in psychotic disorders is nonlinear, with the peak deleterious effects of psychosis on long-term outcomes occurring within the first two years, so that this period should be the focus for intervention[6]. These ideas have been highly influential, with the development of early intervention services explicitly aimed at reducing the DUP[7-9]. To assess how interest in the concept of DUP has developed, we conducted a search of PubMed on July 31, 2020 using the term "duration of untreated psychosis". The results are presented in Figure 1, which shows in-creasing research interest, particularly in the last ten years.

Many mental health services devote significant resources to early intervention in psychosis based, at least partly, on the premise that reducing DUP improves outcomes[10-15]. This notion has been investigated in over a hundred studies examining a number of different outcomes, summarized in several meta-analyses[16-23]. However, due to the inclusion of overlapping samples in different meta-analyses, and differences in inclusion criteria, definition of outcomes, reporting standards and analysis techniques, it is difficult to generate a clear hierarchy of evidence. Furthermore, analyses at first presentation (during the first psychotic episode, soon after the onset of the disease, or at first contact with specialist services) have all included mixed samples of antipsychotic naïve and treated participants[16-18,22]. Thus, no previous analyses have delineated the effects of DUP on outcomes in antipsychotic naïve subjects.

In view of this, we conducted an umbrella review of previous available meta-analyses and performed a new random-effects meta-analysis of primary studies, in order to generate a hierarchical classification of evidence to inform the planning and targeting of interventions to reduce DUP. We aimed to address two related questions: a) what is the relationship between DUP and clinical measures at first presentation?; b) what is the relationship between DUP and outcomes following treatment for psychosis?
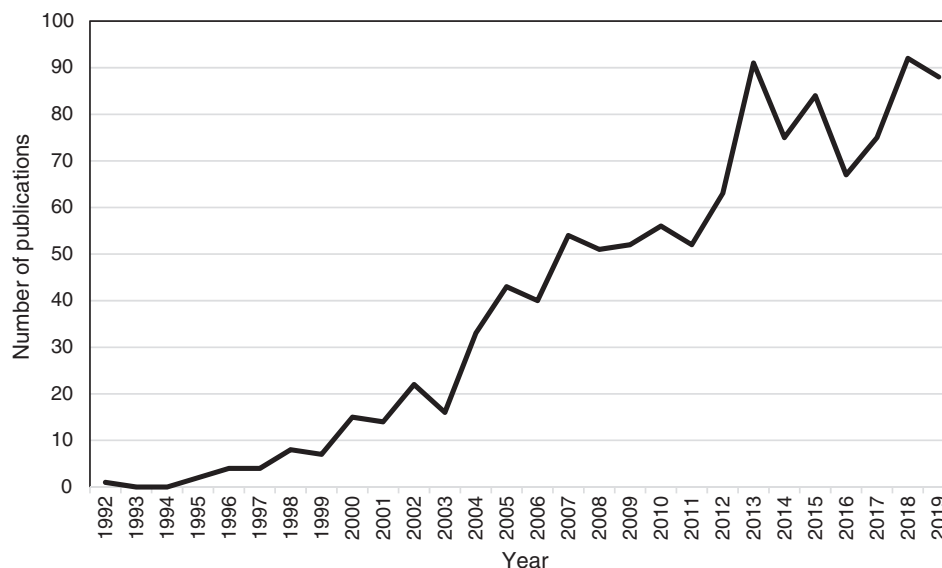
Figure 1 Number of publications per year in PubMed for "duration of untreated psychosis"

METHODS

The umbrella review was performed in line with the relevant guidelines, the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) recommendations, and the Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines[24-26]. The protocol was registered with PROSPERO on June 30, 2020 (no. CRD42020193673) and accepted on August 30, 2020.

Study selection

A search of MEDLINE, Web of Science, PsycINFO and EMBASE was carried out from inception to September 3, 2020, with no date or language restrictions, to identify meta-analyses of studies on the relationship between DUP and outcomes.

We included meta-analyses of studies on patients with schizophrenia spectrum disorders, first-episode psychosis, or affective and non-affective psychosis, which provided data sufficient to allow the calculation of an effect size for the relationship between DUP and outcomes comparable with other studies.

We excluded meta-analyses which: a) focused specifically on affective disorders without psychosis, substance induced psychosis, or psychosis secondary to an organic condition; b) calculated the relationship between DUP and outcomes without using subject level data (e.g., by meta-regression of study level statistics), because this provides a measure of the relationship between DUP and outcomes across studies that is not necessarily comparable with the effect within studies, due to aggregation bias[27-29]. We excluded primary studies which: a) used affective and negative symptoms in the definition of DUP; b) only reported relationships with duration of untreated illness (DUI); c)

were follow-up studies from the pre-antipsychotic era, or studies examining the natural course of psychotic disorders where no subjects were treated; d) were based on carer or patient rated symptom outcomes; e) had more than 10% of participants with substance induced or organic psychosis. Exclusion criteria for primary studies were added after registration of the study protocol, as studies with such designs were not comparable with other included studies, and it was not anticipated that these designs would be encountered.

Only meta-analyses available in English were included, as no systematic bias has been found in meta-analyses including only English language studies, and the majority of countries with specialist early intervention services where DUP research is expected to originate from are either English speaking or have samples which have been extensively described in English[30]. If a primary study was not in English, but was included in a meta-analysis published in English, we included it if sufficient data for analysis and assessment against inclusion criteria could be obtained from the paper and the meta-analysis.

The process from screening to inclusion was conducted independently by two of the authors (MO and LT), with disagreements resolved by discussion. The search strategy used the key words ("systematic review" OR "meta-analysis") AND ("DUP" OR "duration of untreated psychosis" OR "untreated psychosis" OR "duration of untreated illness"). The reference lists of all included papers were also screened to identify further meta-analyses for inclusion. Two authors (MO and TW) individually checked all included meta-analyses and primary studies independently, to assess for overlapping samples, determine selection of outcomes, and ensure that the primary studies met inclusion criteria, resolving disagreements by discussion.

We selected primary studies from identified meta-analyses for two syntheses. To address our first question, we examined

the relationship between DUP and clinical variables at first presentation. To address our second question, we examined the relationship between DUP and outcomes at follow-up following initiation of treatment. Samples could appear in both of these separate analyses if relevant data for each question were available. Cross-sectional studies were considered as part of the first presentation analysis if they took place during the first psychotic episode. Follow-up samples were those assessed after the first psychotic episode or after study baseline in longitudinal studies, regardless of duration of follow-up. Where the information reported was insufficient, corresponding authors were contacted and invited to provide further details.

We identified overlapping samples as recommended in the Cochrane Handbook[31]. When there was substantial overlap, we preferred samples identified from meta-analyses based on individual patient data, if these were available. Otherwise, only the largest dataset was retained for analysis. If the overlap was less than 5%, both samples were included. For follow-up studies, if data for the same sample were available at multiple follow-up points, we preferred the larger sample to maximize sample size, unless the sample sizes were within 15% of each other, in which case we preferred the longer follow-up sample.

Data extraction

From each meta-analysis, we extracted data related to quality of studies, and assessed this quality using the AMSTAR 2 checklist modified for observational studies by Hildebrand et al[32].

All data were re-extracted from each primary study onto piloted forms. One author (MO) performed the primary data extraction, and this was independently checked by at least one other author. Disagreements were resolved by discussion.

From each primary study, we extracted the following data: design, years and location, sample size, patient characteristics, outcomes considered and measures, method of measuring DUP, mean/SD/median of DUP, mean/SD of each outcome (for continuous outcomes), statistics used for analysis (including whether transformations or dichotomization were performed), and effect size. If duration of psychosis was not reported, we calculated it as age at study entry minus age at onset of psychosis.

Where results were presented for pooled outcomes and subgroups, we preferred pooled outcomes to maximize sample size. If results were presented across several different measures on the same outcome (e.g., a study using more than one scale for neurocognition, quality of life or symptoms), effect sizes were averaged across all reported assessment measures, to avoid bias associated with selective preference for significant results.

We preferred unadjusted to adjusted relationships if both were available, as, although adjusted relationships address the issue of confounding, there was no consistency among studies in the variables used for adjustment. If only adjusted relationships were available, we extracted these and planned a sensitivity analysis to exclude such studies. Where data were only available in graphical format, we used WebPlot digitizer to extract them[33].

We included all outcomes considered in the original meta-analyses. As there is no consensus on how to measure some outcomes in psychosis (e.g., remission, quality of life, overall functioning, cognition), we analyzed outcomes as defined in the original meta-analyses, and did not combine similar variables if they were analyzed separately by the original reviews, with the following pre-specified exceptions: a) if the relationship between DUP and outcome was pooled across first presentation and follow-up studies, we separated them to perform two separate analyses; b) if separating the outcomes or including subgroups would allow pooling across different meta-analyses which considered the outcomes separately, we separated outcomes or included the subgroups to maximize overall sample size and ensure consistency in outcome definitions; c) if positive and/or negative symptoms subscale ratings were available, we included these separately, as the relationship between DUP and these outcomes is of clinical interest.

Statistical analysis

We planned to analyze relationships using the effect size measure most commonly reported in the original meta-analysis. However, it became necessary to deviate from the pre-specified protocol, as previous syntheses had combined outcome measures and effect sizes which are not comparable for meta-analysis.

DUP is usually right skewed, as the majority of individuals are treated relatively quickly, but a long tail of people experience a prolonged DUP. For example, in a meta-analysis[34] including 1,391 patients, DUP was not normally distributed, with a mean value (61.7 weeks) exceeding the value of the third quartile (56 weeks), due to the long tail which extended up to 1,200 weeks (23 years). DUP therefore violates the major assumption of the Pearson's product moment correlation, which is that the data are sampled from an underlying bivariate normal distribution[35]. Some of the primary studies of DUP use Pearson's correlation for analyses despite the violation of that assumption, whilst the others use different statistical approaches, either transforming DUP (often with a log or log10 transformation), dichotomizing it into long and short categories, or using non-parametric statistics such as the Spearman's rank correlation coefficient.

The skewed distribution of DUP and these manipulations of the data have important implications for meta-analysis[36], which have not been considered by the majority of previous analyses. Meta-analysis of Pearson's correlations is likely to result in reduced power, and lead to poor performance for point and interval estimates[36,37]. Spearman's and Pearson's correlations should not be combined in the same meta-analysis[38-40]. Effect sizes based on log transformed data should not be combined with untransformed effect sizes in the same meta-analysis[31]. Dichotomization may lead to loss of power and obscure the true relationship between DUP and continuous outcomes, particularly for those with a very long DUP[41]. Moreover, there is no consensus on the threshold separating short vs. long DUP, and cut-off points ranging from four weeks to five years were used in primary studies included in this review[42,43].

The point biserial correlation explores the relationship between a continuous and a dichotomous variable[44]. This correlation may be encountered when studies dichotomize DUP into long/short, or if DUP remains continuous and the outcome is either naturally dichotomous (completed suicide) or artificially dichotomous (high/low symptom scores). When utilizing commonly recommended formulae for converting among effect sizes, it is the point biserial correlation and not the Pearson's correlation which is obtained when converting from means/SDs, t values or Cohen's d into the r family of effect sizes[45]. On the other hand, the phi coefficient represents the correlation between two dichotomous outcomes[44], generated when DUP has been dichotomized and the outcome is either artificially or naturally dichotomous. When utilizing common formulae for converting chi squared statistics into correlations, it is the phi correlation which is obtained[46]. The point biserial correlation and phi correlation coefficients obtained by converting from artificially dichotomized data are not comparable with Pearson's product moment correlations, and should not be combined in the same meta-analyses[44]. Another index, the biserial correlation coefficient, estimates the underlying continuous relationship between a continuous variable and an artificially dichotomized one, and can be synthesized with the Pearson's product moment correlation for the purposes of meta-analysis[44]. The point biserial correlation calculated from artificially dichotomized data is always less than 80% of the biserial correlation[47].

To address all these issues, we analyzed continuous outcomes by using the formulae proposed by Souverein et al[48] to convert Pearson's correlations, Spearman's rank correlations, log transformed correlations, and regression beta values into a single comparable effect size measure, the regression coefficient between log DUP and the log outcome (LogBetaXY). We also calculated the sampling variance of LogBetaXY as recommended by Souverein et al[48]. This approach required the mean and SD of both DUP and the outcome to be reported. If means and SDs were reported separately by subgroup, we calculated the pooled mean/SD using standard formulae[31]. If ranges, medians or interquartile ranges were reported instead of means/SDs, we used Souverein et al's formulae to estimate the log mean and log SD[48]. If no data regarding the mean or SD were reported, we imputed these data, referring to other publications describing the same sample, or, if not available, using results from similar studies. If no comparable data were available for imputation, we excluded these studies.

The above approach assumes that the natural logarithm of DUP and the natural logarithm of the outcome have a bivariate normal distribution, which allows the relationship between DUP and the outcome in the natural scale to be linear, monotonic convex or monotonic concave[48]. Support for this distributional assumption comes from a meta-analysis[18] demonstrating a monotonic concave relationship between DUP and negative symptoms, and a primary study[49] documenting a similar relationship between DUP and Positive and Negative Syndrome Scale (PANSS) total and subscale scores. Due to the double log transformation, the effect size measure (beta) represents the

difference in the log e-transformed predicted value of the outcome for each one-unit difference in the log e-transformed value in DUP[50]. Therefore, an overall beta of 0.1 means that, for every doubling in DUP, the predicted difference in the outcome is 2^{beta} ($2^{0.1}$=1.07 or 7%)[51].

When DUP or the outcome were artificially dichotomized, we employed Jacobs and Viechtbauer's formulae[44] to obtain the biserial correlation coefficient. This coefficient was then used to calculate an estimate of LogBetaXY using the above-mentioned formulae. Its sampling variance was estimated by rearranging the formulae for the sampling variance of LogBetaXY with the Soper's approximate method for the sampling variance of the biserial correlation coefficient described by Jacobs and Viechtbauer[44]. The sampling variance obtained from a biserial correlation is larger than one obtained from a product moment correlation, reflecting the underlying uncertainty associated with the conversion[44].

All calculations were performed in Microsoft Excel, Version 16. All continuous data are expressed such that a negative value indicates a relationship between DUP and poorer outcome (for example, more severe symptoms, poorer functional status, smaller reduction in symptoms).

For categorical outcomes, we synthesized effect size measures using the odds ratio (OR). If the point biserial correlation was reported, we calculated the OR using standard formulae[45]. Where 2x2 tables were reported or could be constructed, we calculated the OR and its sampling variance using standard formulae[52]. When means/SDs of DUP were reported at the level of the dichotomous outcome, we calculated the Cohen's d effect size and then converted it to the (log) OR and its standard deviation using standard formulae[45]. For the few studies that reported hazard ratios, we estimated the OR using previously proposed formulae[53]. All categorical data are presented such that an OR above 1 indicates a relationship between DUP and poorer outcome.

Final value and change in correlations were not combined in the same analysis, including syntheses of treatment response with remission. Effect size measures for truly binary outcomes, artificially dichotomized outcomes and continuous outcomes were not combined in the same meta-analysis. Log transformed and untransformed effect sizes were not combined in the same meta-analysis. Studies where a comparable effect measure and outcome measure could not be calculated were excluded. We only performed meta-analysis when there were more than three studies.

Random-effects meta-analysis

Data were analyzed with the metafor package in R to calculate the random-effects p value, effect size, confidence interval, heterogeneity (I^2) and prediction interval for each outcome[54]. Random-effects models were used as we anticipated considerable heterogeneity in DUP definitions and values, outcome definitions and sample characteristics. Where there were two subsamples from the same study reporting effect sizes, the subsamples

were first combined using fixed effects meta-analysis[31]. If significant relationships were reported only for one subsample or outcome, with no comment on results in the other subsample(s), the other subsample(s) was assumed to have an effect size of 0 to be conservative.

We performed Egger's test for small study effects[55]. A p value <0.10 combined with a more conservative effect in the largest study than in the random-effects meta-analysis was judged to provide evidence for small study effects, as in previous umbrella reviews[56]. When Egger's test was significant, we used the Duval and Tweedie's trim-and-fill procedure to estimate true effects controlling for any detected bias[57].

Excess significance bias was calculated using Ioannidis and Trikalinos' test[58]. With the metaviz package in R, we estimated the power of each study using a non-central p distribution[59]. The sum of all power estimates provides the expected number of significant datasets. The actual observed number of statistically significant datasets is then compared to the expected number using a χ^2-based test. Significance was assessed at two-sided p<0.10 with observed > expected, as in previous umbrella reviews[56].

For significant results, we also conducted "file-drawer" analysis, where we calculated the number of fail-safe studies that would have to be added to the observed set of results to reduce the p value associated with the weighted average random-effects effect size to 0.05[60].

We applied the following criteria to assess the level of evidence for the association between DUP and outcomes, as in previous umbrella reviews[56]: a) convincing (class I): meta-analysis based on sample size >1,000, results show significance with $p<10^{-6}$, $I^2<50\%$, 95% prediction interval excluding the null, no small study effects, and no excess significance bias; b) highly suggestive (class II): N>1,000, $p<10^{-6}$, largest study with a statistically significant effect, and class I criteria not met; c) suggestive (class III): N>1,000, $p<10^{-3}$, and class I-II criteria not met; d) weak (class IV): p<0.05 and class I-III criteria not met; e) non-significant: p>0.05.

Outliers, heterogeneity assessment, meta-regression, sensitivity and subgroup analyses

We used the above software to run analyses with and without outliers, defined as studies whose effect size confidence interval did not overlap with the confidence interval of the pooled effect size[61]. We calculated I^2 and Cochrane's Q to test for heterogeneity of study effects.

Meta-regression required a minimum of ten complete data points for continuous variables, and four studies per subgroup for categorical variables, to ensure adequate power[29]. The p values for meta-regression were corrected using the Benjamini-Hochberg procedure, with a false discovery rate (FDR) of 5%[62].

For the purposes of meta-regression, DUP startpoint, DUP endpoint and previous antipsychotic exposure were assigned into categories to see if these moderated effects. Samples were categorized into antipsychotic naïve (all participants antipsychotic naïve at study entry), minimal antipsychotic treatment

(all participants had received less than 1-month antipsychotic treatment, or more than 75% participants were antipsychotic naïve and the others had less than 3-month antipsychotic exposure at study entry), and appreciable antipsychotic treatment (greater than 1-month antipsychotic treatment at study entry, or first presentation measures recorded at or after end of first hospitalization)[63]. If the duration of previous treatment was unclear, samples were categorized in the appreciable antipsychotic treatment group if the majority of participants had been exposed to antipsychotics, and in the minimal group if the majority were antipsychotic naïve. Studies in which previous antipsychotic exposure was unclear were excluded from this analysis.

DUP onset definitions and DUP endpoints varied among studies. We adapted the criteria used by Oliver et al[64] to define DUP onset as either the onset of the first ever recalled psychotic symptom, or the point at which psychotic symptoms met a clearly defined threshold (either above a cut-off on the PANSS, a description of "clear" or "overt" psychotic symptoms, or continuous psychotic symptoms over a given time period). We did not distinguish between different symptom, severity or duration thresholds used. DUP endpoints were categorized as initiation of antipsychotic treatment, first hospitalization, first contact with health services or study entry, and endpoints requiring either a response to treatment or a specified duration of treatment (such as 4-week antipsychotic treatment).

Meta-regression was undertaken after the removal of outlying studies, as defined above. This analysis aimed to test if there was a relationship between year of publication, scale used to assess outcome measure, duration of follow-up and dropout percentage (for follow-up studies), percent of subjects diagnosed with schizophrenia (or, separately, percent diagnosed with schizophrenia spectrum disorders if insufficient studies reported the percent with schizophrenia), mean age, mean duration of psychosis, gender composition, mean DUP, DUP startpoint, DUP endpoint, and statistics used to calculate effect size.

Where sample sizes permitted, we performed subgroup analyses on subjects who were antipsychotic naïve at study entry (in first presentation analyses) and on studies excluding patients with affective psychosis. We performed planned sensitivity analyses removing studies that provided adjusted relationships, those where data were imputed from other samples, and those including any participants with drug induced or organic psychosis. For follow-up studies, subgroup analysis was conducted on prospective studies only for variables rated class I to III.

RESULTS

Included studies

The systematic search identified 149 unique meta-analyses (Figure 2). Two additional items were identified through being referenced in the included papers. Of these, thirteen meta-analyses met inclusion criteria. A full list of excluded studies with reasons for exclusion is provided in the supplementary information.
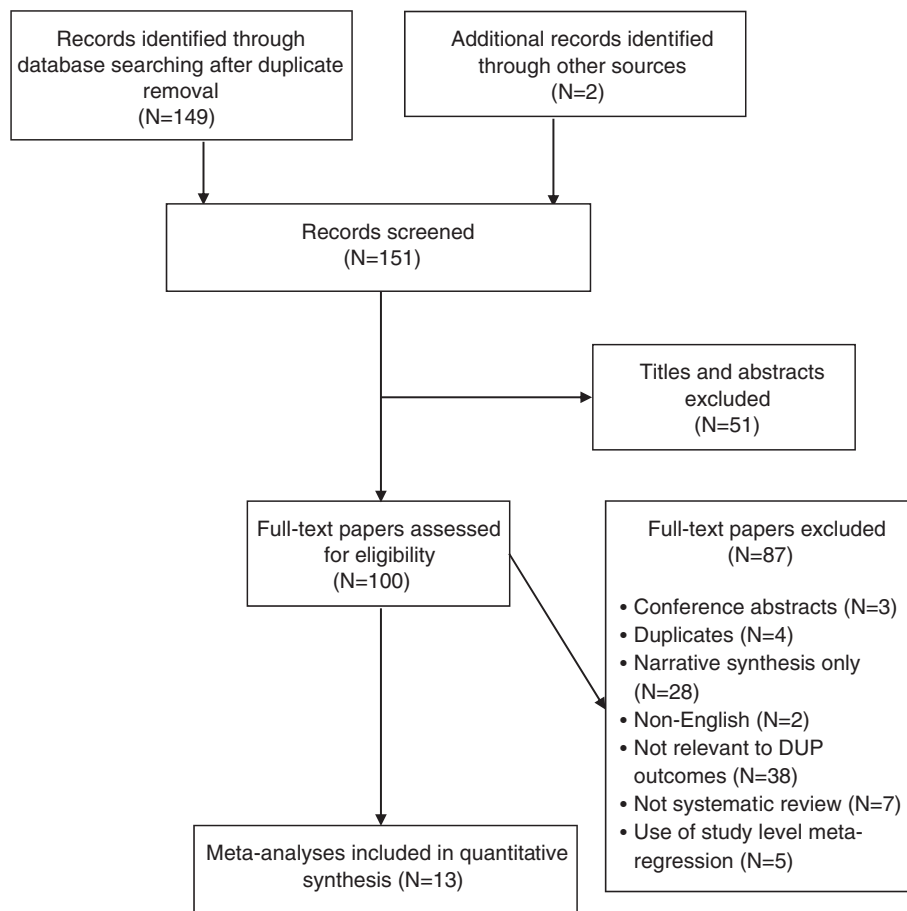
Figure 2 PRISMA flow chart. DUP – duration of untreated psychosis

Table 1 summarizes the meta-analyses included, with their flaws and other methodological considerations. From these meta-analyses, we identified 129 reports of non-overlapping primary studies for inclusion, with a total sample size of 25,657 patients. Some studies appeared in multiple meta-analyses; they were coded as being identified from the most recent meta-analysis. The list of the primary studies included is provided in the supplementary information.

Definitions of outcomes

As pre-specified, we avoided redefining outcomes as much as possible. However, there were discrepancies between meta-analyses on definitions of some outcomes, and some meta-analyses combined effect measures and outcomes which were not comparable. We defined overall, social and vocational functioning as in Santesteban-Echarri et al[20], relapse as in Alvarez-Jimenez et al[67], global psychopathology as in Perkins et al[17], and remission as in Marshall et al[16]. We conducted subgroup analysis of studies which defined remission as in Penttila et al[19], using the operationalized Andreasen et al's consensus criteria[70]. We

combined violence and serious violence into one category since, after excluding one study on serious violence which measured DUI, there were only two remaining studies assessing serious violence, and both were subgroup analyses in studies also assessing violence.

Hospitalization was the only outcome not defined as in any previous meta-analyses. Some studies which were included in "hospital treatments" in Penttila et al[19] were re-categorized as assessing relapse for consistency with Alvarez-Jimenez et al[67], and the remaining studies measured either duration of hospitalization or number of hospitalizations. We considered these two outcomes separately as they measured different underlying constructs.

Relationship between DUP and clinical variables at first presentation

The relationship between DUP and clinical variables at first presentation is summarized in Table 2, and in Figure 3 for continuous variables and Figure 4 for categorical variables. At first presentation, there was suggestive (class III) evidence for a rela-

tionship between longer DUP and more severe negative symptoms and greater risk of previous self-harm, and weak (class IV) evidence for a relationship between longer DUP and poorer quality of life. There was no significant relationship between DUP and positive symptoms, global cognition, overall functioning, global psychopathology, risk of violence, and cannabis, alcohol or substance misuse at first presentation.

There was evidence of significant publication bias and small study effects for negative symptoms (Egger's test p=0.045). There was no evidence of significant publication bias, excess significance bias or small study effects for the other significant variables (Egger's test p=0.24 for deliberate self-harm, p=0.49 for quality of life). Using the trim-and-fill method, no studies were imputed on the right-hand side for negative symptoms. "File-drawer" analysis suggested that the significant results for deliberate self-harm and negative symptoms would require 30 and 559 missing studies, respectively, with an effect size of 0 to negate their statistical significance. The overall random-effects result for the quality of life analysis was marginally significant and, accordingly, only one study would be required to negate its significance.

There was no statistically significant evidence of heterogeneity in analyses of cannabis misuse, alcohol and substance misuse, global cognition, deliberate self-harm, or overall functional status. We encountered substantial heterogeneity in our analyses of negative symptoms, quality of life, violence, global psychopathology and positive symptoms (all p values <0.0001) (see Table 2).

Removal of outliers led to variable reductions in heterogeneity, causing absolute reductions in I^2 between 17 and 30% for negative symptoms, quality of life and positive symptoms, with a 3% reduction seen in global psychopathology. No statistically significant result changed from significant to non-significant after removal of outliers. On the contrary, all classes of evidence remained the same with the exception of quality of life, which increased from class IV to class III due to a decrease in the random-effects p value.

Meta-regression was conducted to explore the residual heterogeneity in the relationship between DUP and negative symptoms, quality of life, and positive symptoms. For negative symptoms, year of publication and DUP endpoint definition were significant predictors after Benjamini-Hochberg correction, with a FDR corrected p value of 0.039 and 0.001, respectively. Studies which were published more recently reported a smaller relationship between DUP and negative symptoms (intercept=–12.5845, beta=0.0064, residual I^2=59%). Studies which used hospitalization as the endpoint for DUP reported a larger effect size for the relationship between DUP and negative symptoms (beta=–0.11) compared to those which used adequate treatment (beta=–0.02) or initiation of treatment (beta=–0.05). There was no significant residual heterogeneity in the negative symptom analysis (I^2=27%, p>0.05) after inclusion of DUP endpoint in the random-effects model.

Using meta-regression, we did not find any moderator variables to explain the remaining heterogeneity following removal of outliers in quality of life (I^2=77%, p=0.0001) or positive symp-

toms (I^2=59%, p=0.02). These analyses were limited, as we were only able to examine the effect of three moderator variables for positive symptoms and one for quality of life, due to sample size limitations. Although there was also substantial unexplained heterogeneity in the violence analysis (I^2=85%), there were no outliers and too few data points for meta-regression for this variable and for global psychopathology (I^2=81% after removal of outliers).

For the majority of analyses, sensitivity analysis which excluded samples recruiting participants with affective psychosis had no discernible impact on the heterogeneity. The exceptions were alcohol and substance misuse, where I^2 dropped from 32% to 0%, and deliberate self-harm, where I^2 increased from 0% to 56%. Removing studies that included patients with affective psychosis also did not affect the class of evidence for most analyses. However, in the negative symptom analysis, removing the eight samples which included participants with affective psychosis reduced the class of evidence from III to IV, due to an increase in the random-effects p value from 3.6×10^{-5} to 0.003. For deliberate self-harm and quality of life, removing these samples reduced the class of evidence from III and IV respectively to non-significant, because the random-effects p value became >0.05.

There was also no discernible impact on the heterogeneity when we removed the small number of samples which included participants with drug induced psychosis (up to 10%) from the negative symptoms, quality of life, deliberate self-harm, global cognition, violence and substance misuse analyses, apart from finding that the relationship between DUP and quality of life decreased from class IV to non-significant, because the random-effects p value became 0.10. Inclusion of adjusted effect sizes and imputations of the mean/SDs of DUP and/or the outcome from other samples had no effects on classes of evidence and minimal effect on heterogeneity for all analyses.

We conducted subgroup analyses of antipsychotic naïve subjects where data were available. We found that there was an absolute reduction in I^2 of 23% for the relationship between DUP and negative symptoms after removal of patients who had received any previous antipsychotic treatment, and results remained statistically significant.

Relationship between DUP and outcomes at follow-up

The relationship between DUP and outcomes at follow-up is summarized in Table 3, and in Figure 5 for continuous outcomes and Figure 6 for categorical outcomes.

We found highly suggestive (class II) evidence for a relationship between longer DUP and more severe negative symptoms, more severe positive symptoms and lower chance of remission at follow-up. We found suggestive (class III) evidence for a relationship between longer DUP and more severe global psychopathology and poorer overall functional outcome at follow-up. There was weak (class IV) evidence for a relationship between longer DUP and poorer social and vocational functioning, poorer quality of life, and smaller reduction in total symptoms at follow-up.

Table 1 Description of included meta-analyses

	Variables assessed at first presentation and follow-up	AMSTAR 2	Other methodological comments
Marshall et al[16]	First presentation None Follow-up (prospective, 2 months - 1 year) Remission	Total: 4/14 6/7 critical flaws relating to: • Lack of pre-registration • No reporting of excluded studies and reasons for exclusion • Lack of satisfactory technique for assessing risk of bias in included studies • Statistical techniques – Lack of adjustment for heterogeneity • No assessment of publication bias • No discussion of impact of risk of bias on results	Includes one sample of "naturalistic" never-treated patients from pre-AP era in a follow-up comparison with long-term AP-treated patients. Combines change in and final value positive symptom scores. Includes one study with affective and negative symptoms in definition of DUP. Combines Pearson's correlation coefficients with Spearman's rank correlations, and with Pearson's correlation between log transformed DUP and outcome on natural scale. However, conducts sensitivity analyses excluding the untransformed Pearson's correlations.
Perkins et al[17]	First presentation Global psychopathology, positive symptoms, negative symptoms, overall functional status Follow-up (prospective, 1 month - 15 years) Global psychopathology, positive symptoms, negative symptoms, overall functional outcome, remission	Total: 2/14 7/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not search at least 2 databases, did not justify search restrictions, did not search trial/study registries) • No reporting of excluded studies and reasons for exclusion • No assessment of risk of bias in included studies • Statistical techniques – Lack of justification for unadjusted effects and no adjustment for heterogeneity • No assessment of publication bias • No discussion of impact of risk of bias on results	Includes one study with affective and negative symptoms in definition of DUP. Includes one sample which assessed DUI, not DUP. Combines change in and final value positive symptom scores. Combines Pearson's correlation coefficients with Spearman's rank correlations, and with Pearson's correlation between log transformed DUP and outcome on natural scale.
Farooq et al[65]	First presentation Positive symptoms Follow-up (prospective, 1 month - 2 years) Reduction in total symptoms, overall functional outcome	Total: 4/14 7/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not search trial/study registries) • No reporting of excluded studies and reasons for exclusion • No assessment of risk of bias in included studies • Statistical techniques – Lack of investigation of heterogeneity • No assessment of publication bias • No discussion of impact of risk of bias on results	Includes one sample of "naturalistic" never-treated patients in a follow-up comparison with long-term AP-treated patients. Combines outcome measure of remission in these patients with scale-assessed overall function in other samples. One significant data extraction error: uses SE of beta value as correlation coefficient. Combines adjusted and unadjusted effect size measures with no sensitivity analysis. Combines Pearson's correlation coefficient with phi coefficient, point biserial coefficient, and Spearman's rank correlation.
Large & Nielssen[66]	First presentation Violence/serious violence Follow-up None	Total: 5/14 6/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify search restrictions, did not search trial/study registries) • No reporting of excluded studies and reasons for exclusion • No assessment of risk of bias in included studies • Statistical techniques – Lack of investigation of heterogeneity and unclear if authors used adjusted or unadjusted effects with lack of justification for choice • No discussion of impact of risk of bias on results	Combines log transformed and untransformed effect sizes. Unit of analysis error whereby the same sample is included twice in the same meta-analysis for different outcomes, causing it to be given disproportionate weight. Assessment time frames for violence vary, encompassing lifetime violence, violence only during DUP, violence preceding/precipitating first admission and violence during defined time periods prior to assessment. Includes one sample which assessed DUI, not DUP.

Table 1 Description of included meta-analyses (continued)

	Variables assessed at first presentation and follow-up	AMSTAR 2	Other methodological comments
Alvarez-Jimenez et al[67]	First presentation None Follow-up (prospective, 1-7.5 years) Relapse	Total: 7/14 4/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not provide search strategy) • No reporting of excluded studies and reasons for exclusion • No discussion of impact of risk of bias on results	Varying definitions of relapse: readmission to hospital due to psychosis, symptom defined and combinations of the two.
Boonstra et al[18]	First presentation Negative symptoms Follow-up (prospective, 1-8 years) Negative symptoms	Total: 4/14 6/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify publication restrictions, did not search trial/study registries and reference lists of included studies, did not conduct search within 24 months of completion of review) • No reporting of excluded studies and reasons for exclusion • No assessment of risk of bias in included studies • Statistical techniques – Lack of adjustment for heterogeneity • No discussion of impact of risk of bias on results	Obtains individual patient data for each study in order to compute comparable summary effect sizes for all (Spearman's rank correlation).
Burns[68]	First presentation Alcohol and substance misuse, cannabis use Follow-up None	Total: 3/14 6/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify search restrictions, did not search trial/study registries) • No reporting of excluded studies and reasons for exclusion • Statistical techniques – Lack of adjustment for heterogeneity • No assessment of risk of bias in included studies • No discussion of impact of risk of bias on results	Includes three overlapping samples on the same outcome. Assessment time frames for substance misuse vary, encompassing either lifetime misuse, a diagnosable substance misuse disorder at the time of assessment, or misuse during a previous defined time period. ~60% of studies assess alcohol and substance misuse, ~40% assess substance misuse alone. Combines adjusted and unadjusted effect sizes with no sensitivity analysis.
Challis et al[69]	First presentation Deliberate self-harm Follow-up (prospective and retrospective, 1.5-4 years) Deliberate self-harm	Total: 8/14 5/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify search restrictions, did not search trial/study registries) • Lack of satisfactory technique for assessing risk of bias in included studies • Statistical techniques – Unclear if authors used adjusted or unadjusted effects, lack of justification for choice • Lack of satisfactory technique for assessing risk of bias	Combines log transformed and untransformed effect sizes. Combines adjusted and unadjusted effect sizes with no sensitivity analysis. Assessment time frames for DSH vary, encompassing either lifetime self-harm, self-harm during DUP, DSH during a previous defined time period, and unspecified.

Table 1 Description of included meta-analyses (*continued*)

	Variables assessed at first presentation and follow-up	AMSTAR 2	Other methodological comments
Penttila et al[19]	First presentation None Follow-up (prospective, cross-sectional and retrospective, 2-28 years) Global psychopathology, positive symptoms, negative symptoms, number of hospitalizations, time hospitalized, relapse, overall functional outcome, quality of life, remission, social functioning, vocational functioning	Total: 6.5/14 5/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not search trial/study registries and reference list of included studies) • Did not report full list of excluded studies and reasons for exclusion • Did not use a satisfactory method for assessing risk of bias in included studies • Statistical techniques – Lack of adjustment for heterogeneity	Minor data extraction errors. One outcome (vocational functioning) included in social functioning meta-analysis in error. Includes one study with carer-rated symptoms. Includes one study with negative symptoms and social decline in DUP definition. Combines relationship between DUP and level of final value negative symptoms with point biserial correlation between DUP and persistent negative symptoms. Combines relationship between DUP and change in symptom scores with relationship between DUP and final value symptom scores. GAF appears as a measure of social functioning and global outcome in different studies. Combines Pearson's correlation coefficients with phi correlation, point biserial correlation, Spearman's correlations, Pearson's correlation between log transformed DUP (log 10 and/or natural log) and outcome in natural scale, and Pearson's correlation between log transformed DUP and the log transformed outcome. All outcome categories other than quality of life include studies measuring different underlying constructs.
Santesteban-Echar-ri et al[20]	First presentation None Follow-up (prospective, 1-5 years) Overall functional outcome, quality of life, social functioning	Total: 7/14 4/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify language restrictions) • Lack of satisfactory technique for assessing risk of bias • Did not satisfactorily discuss impact of risk of bias on results	Three sets of overlapping samples included in same meta-analysis (7 studies in total). Combines Pearson's correlation coefficients with phi correlation, point biserial correlation, Pearson's correlation between log transformed DUP (log 10 and/or natural log) and the outcome in its natural scale, and Pearson's correlation between log transformed DUP and log transformed outcome.
Allott et al[21]	First presentation Global cognition Follow-up None	Total: 7/14 5/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify language restrictions, did not search reference lists of all included studies, did not search trial registries) • Did not report a full list of excluded studies and reasons for exclusion • Statistical techniques – Does not state whether used adjusted or unadjusted effect sizes and does not justify this choice • Did not discuss impact of risk of bias on results	Combines unmedicated and AP-treated patients in same meta-analysis. Combines adjusted and unadjusted effect sizes with no sensitivity analysis. Minor data extraction errors. Combines Pearson's correlation coefficients with point biserial correlation, Spearman's correlations, and Pearson's correlation between log transformed DUP (log 10 and/or natural log) and outcome in natural scale.

Table 1 Description of included meta-analyses (*continued*)

	Variables assessed at first presentation and follow-up	AMSTAR 2	Other methodological comments
Bora et al[22]	First presentation Global cognition Follow-up (prospective, 3 months-3 years) Global cognition	Total: 6.5/14 5/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not justify search restrictions, did not search trial/study registries) • Did not report a full list of excluded studies with reasons for exclusion • Statistical techniques – Does not state whether used adjusted or unadjusted effect sizes and does not justify this choice, does not fully investigate heterogeneity • Did not discuss impact of risk of bias on results	Combines unmedicated and AP-treated patients in same meta-analysis. Combines adjusted and unadjusted effect sizes with no sensitivity analysis. Combines Pearson's correlation coefficients with point biserial correlation, Spearman's correlations, and Pearson's correlation between log transformed DUP (log 10 and/or natural log) and outcome in natural scale.
Watson et al[23]	First presentation Quality of life Follow-up (prospective, 6 months-12 years) Quality of life	Total: 4/14 6/7 critical flaws relating to: • Lack of pre-registration • Systematic search (did not search at least two databases, justify publication restrictions, search trial/study registries, search reference lists of included studies) • Did not report a full list of excluded studies and reasons for exclusion • Statistical techniques – Lack of justification for unadjusted effects and no adjustment for heterogeneity • No assessment of risk of bias in included studies • No discussion of impact of risk of bias on results	Combines first-presentation, never-medicated patients with long-term follow-up (8 and 12 years) in same meta-analysis. No presentation of effect sizes or sampling variance for meta-analysis. Unclear which measure was chosen when results for different time points and subscales were available. Combines Pearson's correlation with Spearman's correlation coefficient, and Pearson's correlation between log transformed DUP (log 10 and/or natural log) and outcome in natural scale.

DUP – duration of untreated psychosis, AP – antipsychotic, DSH – deliberate self-harm, DUI – duration of untreated illness, GAF – Global Assessment of Functioning, SE – standard error

Table 2 Evidence for associations between duration of untreated psychosis (DUP) and clinical variables at first presentation

Outcome	Studies	Total sample size (N)	Random-effects p value	Random-effects measure (95% CI)	I² (p value)	95% prediction interval	Small study effects/excess significance bias	Largest study significant	Class of evidence	Predicted difference in continuous outcome for every doubling in DUP
Negative symptoms	23	4,165	3.60x10⁻⁵	beta=−0.07 (−0.10 to −0.04)	87.1% (<0.0001)	−0.21; 0.08	No/Yes (fail-safe N: 559)	No	III	5% worse
Deliberate self-harm	8	1,752	1.07x10⁻⁵	OR=1.89 (1.42-2.52)	0% (0.15)	1.42; 2.52	No/No (fail-safe N: 30)	Yes	III	NA
Quality of life	9	1,726	0.044	beta=−0.14 (−0.29 to −0.004)	99.2% (<0.0001)	−0.58; 0.29	No/No (fail-safe N: 1)	Yes	IV	11% worse
Global cognition	14	1,970	0.17	beta=−0.01 (−0.02 to 0.004)	1.0% (0.85)	−0.02; 0.005	No/No	Yes	NS	NA
Violence and serious violence	4	1,008	0.23	OR=1.66 (0.70-3.94)	89% (<0.0001)	0.26; 10.48	No/No	No	NS	NA
Alcohol and substance misuse	7	2,281	0.29	OR=0.88 (0.7-1.1)	31.9% (0.25)	0.59; 1.32	No/No	No	NS	NA
Global psychopathology	8	796	0.33	beta=−0.02 (−0.05 to 0.02)	83.5% (<0.0001)	−0.11; 0.07	No/Yes	Yes	NS	NA
Positive symptoms	12	1,084	0.55	beta=0.01 (−0.03 to 0.05)	88.9% (<0.0001)	−0.12; 0.15	No/No	No	NS	NA
Overall functional status	4	333	0.60	beta=−0.04 (−0.21 to 0.12)	0% (0.84)	−0.21; 0.12	No/No	No	NS	NA
Cannabis misuse	7	1,508	0.90	OR=0.99 (0.82-1.19)	0% (0.58)	0.82; 1.19	No/No	No	NS	NA

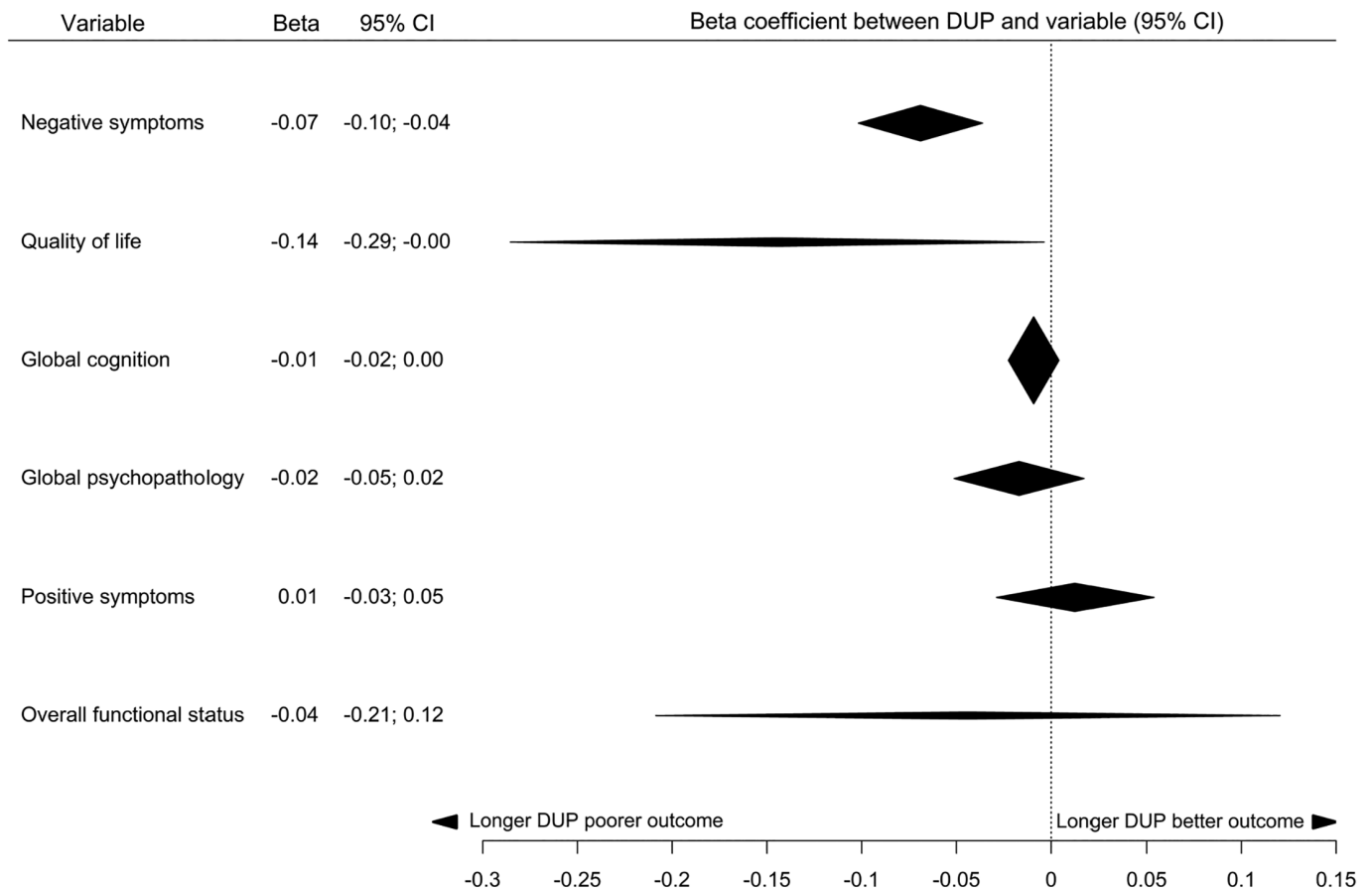OR – odds ratio, NS – not significant, NA – not applicable

Figure 3 Summary of effect sizes for relationships between duration of untreated psychosis (DUP) and continuous clinical variables at first presentation

In follow-up studies, there was no significant relationship between DUP and risk of relapse, risk of deliberate self-harm, global cognition, time hospitalized, and number of hospitalizations.

Egger's test was statistically significant with evidence of small study effects for the analyses of positive symptoms (p=0.025), remission (p<0.001) and number of hospitalizations (p<0.001). Using the trim-and-fill method, no studies were imputed on the right-hand side for positive symptoms or number of hospitalizations. Seven studies were imputed on the left-hand side in the remission analysis; the class of evidence remained unchanged. "File-drawer" analysis showed that more than 1,650 null studies would be needed to nullify the results of the negative symptom analysis, whereas the marginally significant results for vocational functioning, reduction in total symptoms and quality of life would require only one null study.

There was no statistical evidence of heterogeneity in analyses of social functioning, vocational functioning or deliberate self-harm at follow-up. There was mild heterogeneity present in global cognition (p=0.01). We encountered moderate to substantial heterogeneity in negative symptoms, positive symptoms, remission, overall functional outcome, global psychopathology, reduction in total symptoms, quality of life, relapse, and number of hospitalizations (all p<0.0001).

Removal of outliers led to large (21-64%) absolute reductions in I^2 for negative symptoms, relapse, quality of life, overall functional outcome and remission. There were smaller reductions (5-12%) in heterogeneity for positive symptoms and global psychopathology. The majority of results were minimally affected by removal of outliers – no results went from significant to non-significant, although remission decreased from class II to class III, due to removal of the largest significant study, despite a large decrease in the random-effects p value ($3×10^{-9}$ to $2×10^{-19}$). Global psychopathology, overall functional outcome, and quality of life increased class of evidence (from III to II, III to II, and IV to III, respectively) following outlier removal, due to decreases in the random-effects p values.

Where sample sizes allowed, meta-regression was conducted for outcomes with moderate to substantial heterogeneity remaining after outlier removal. There were insufficient data available for exploration of the residual heterogeneity in quality of life, relapse, reduction in total symptoms, global cognition and the hospitalization outcomes. For positive symptoms, no potential moderators survived FDR correction. For negative symptoms, dropout percent (corrected p=0.035) survived FDR correction. Studies where fewer subjects were lost to follow-up (intercept=–0.1364, beta=0.2247, residual I^2=44%) reported

Variable	OR	95% CI	Odds ratio between DUP and variable (95% CI)
Deliberate self-harm	1.89	1.42; 2.52	
Violence/Serious violence	1.66	0.70; 3.94	
Alcohol/Substance misuse	0.88	0.70; 1.11	
Cannabis misuse	0.99	0.82; 1.19	

◄ Longer DUP better outcome   Longer DUP poorer outcome ►

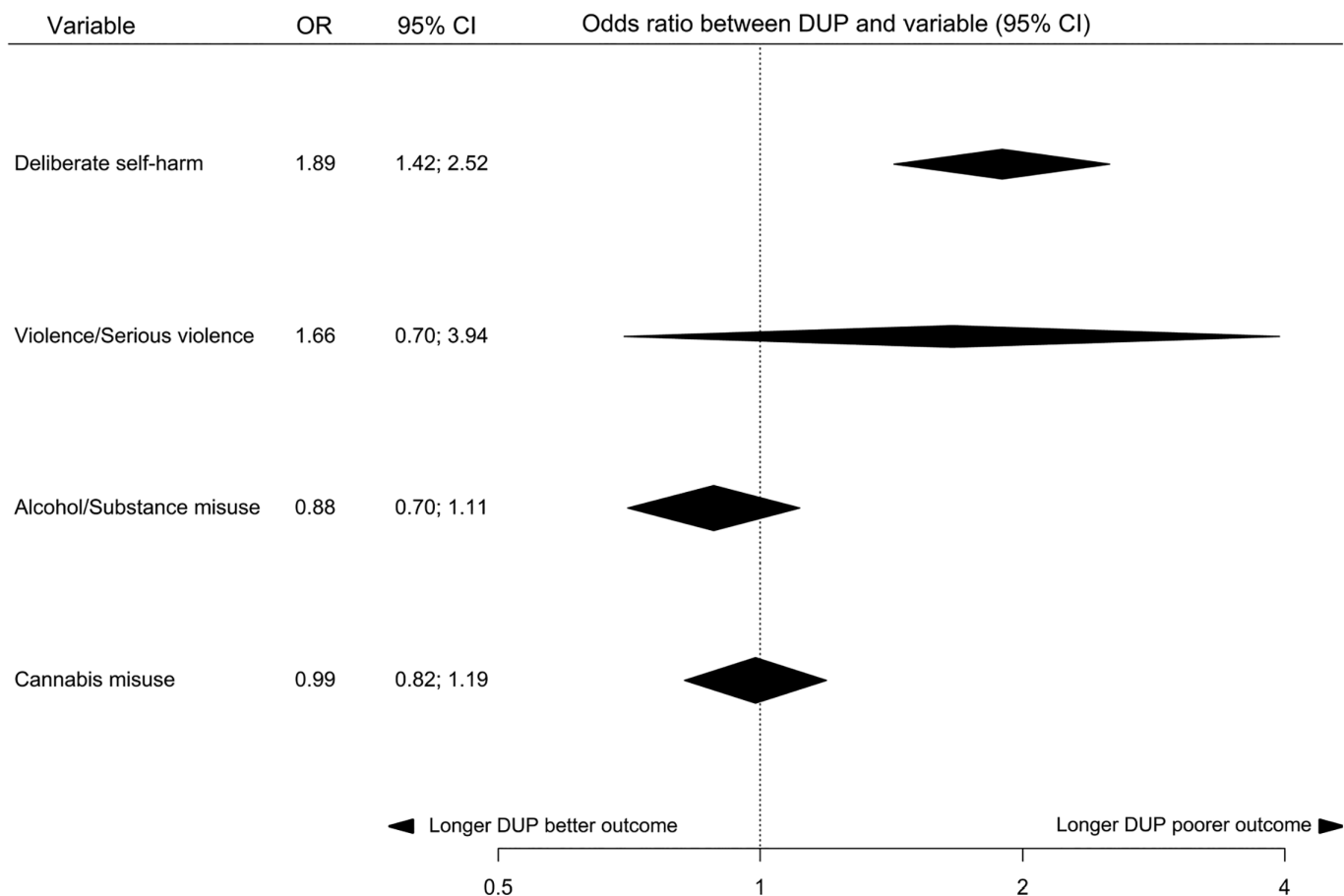0.5   1   2   4

Figure 4 Summary of effect sizes for relationships between duration of untreated psychosis (DUP) and categorical clinical variables at first presentation

larger relationships between DUP and negative symptoms. For global psychopathology, percent of subjects with schizophrenia (corrected p=0.0003) and dropout percent (corrected p=0.044) survived Benjamini-Hochberg correction. Studies with higher proportions of subjects with schizophrenia (intercept=–0.0260, beta=–0.1530, residual I^2=36%) and studies where fewer subjects were lost to follow-up (intercept=–0.1819, beta=0.2658, residual I^2=42%) reported larger relationships between DUP and global psychopathology.

For overall functional outcome, the definition of the endpoint of DUP moderated the effects seen. Studies which used the initiation of antipsychotic treatment as the endpoint for DUP reported larger effects than those using adequate antipsychotic treatment (corrected p=0.022; beta=–0.06 for studies using adequate treatment, beta=–0.11 for studies using the initiation of treatment). There was no statistically significant heterogeneity following inclusion of DUP endpoint definition in the model (I^2=0%, p=0.44).

For the majority of outcomes, sensitivity analysis which excluded samples recruiting participants with affective psychosis had no discernible impact on the heterogeneity. The exceptions were quality of life and remission, where I^2 fell by 51% and 59%,

respectively. For positive symptoms, removing these samples reduced the class of evidence from II to III, through an increase in the random-effects p value from 5x10^{-8} to 4x10^{-5}. There was no effect on the class of evidence for any other analysis. There was one sample which included people with drug induced psychosis in each of the social functioning, remission and overall functioning analyses. Removal of this sample had no discernible impact on results. Restricting analysis of studies examining remission to those using Andreasen et al's operationalized criteria[70] reduced the class of evidence from II to IV, due to an increase in the random-effects p value.

Imputations of the mean/SDs of DUP and/or the outcome from other samples had no effect on the class of evidence and a negligible effect on heterogeneity in most analyses. However, for global psychopathology, removing studies where data were imputed reduced I^2 by 20% and the class of evidence from III to IV, due to a reduction in the sample size below the class III threshold of 1,000, although the p value was more significant. Overall, findings were similar when removing studies which calculated adjusted effect sizes, and most analyses remained in the same class of evidence. The exception was remission, where heterogeneity fell to 0% and the class of evidence decreased from II to III,

Table 3 Evidence for associations between duration of untreated psychosis (DUP) and outcomes at follow-up

Outcome	Studies	Mean/median follow-up (years)	N	Random-effects p value	Random-effects measure (95% CI)	I^2 (p value)	95% prediction interval	Small study effects/excess significance bias	Largest study significant	Class of evidence	Predicted difference in continuous outcome for every doubling in DUP
Negative symptoms	27	5.9/2	3,633	3.46×10^{-10}	beta=-0.11 (-0.15 to -0.08)	86.7% (<0.0001)	-0.27; 0.05	No/No (fail-safe N: 1,667)	Yes	II	8% worse
Remission	22	6.1/2.1	3,570	2.98×10^{-10}	OR=2.16 (1.7-2.75)	63.7% (<0.0001)	1.003; 4.67	Yes/Yes (fail-safe N: 46)	Yes	II	NA
Positive symptoms	21	6.8/3	2,934	4.52×10^{-8}	beta=-0.16 (-0.22 to -0.11)	94.4% (<0.0001)	-0.42; 0.10	Yes/No (fail-safe N: 47)	Yes	II	12% worse
Global psychopathology	14	9.1/10.6	1,412	4.72×10^{-6}	beta=-0.16 (-0.22 to -0.09)	96.0% (<0.0001)	-0.41; 0.09	No/No (fail-safe N: 3)	Yes	III	12% worse
Overall functional outcome	27	6.9/3	3,104	2.16×10^{-6}	beta=-0.11 (-0.16 to -0.07)	96.3% (<0.0001)	-0.35; 0.12	No/No (fail-safe N: 13)	Yes	III	8% worse
Social functioning	4	5.5/4.5	286	1.38×10^{-11}	beta=-0.06 (-0.08 to -0.04)	0% (0.64)	-0.08; -0.04	No/No (fail-safe N: 41)	Yes	IV	4% worse
Vocational functioning	3	9/10	371	0.0005	beta=-0.04 (-0.06 to -0.02)	0% (0.61)	-0.06; -0.02	No/Yes (fail-safe N: 1)	Yes	IV	3% worse
Reduction in total symptoms	4	0.6/0.16	350	0.031	beta=-0.14 (-0.26 to -0.01)	94.5% (<0.0001)	-0.41; 0.14	No/No (fail-safe N: 1)	Yes	IV	10% worse
Quality of life	8	3.7/1.5	1,162	0.042	beta=-0.09 (-0.17 to -0.003)	96.9% (<0.0001)	-0.33; 0.16	No/No (fail-safe N: 1)	Yes	IV	6% worse
Relapse	9	4/2	1,264	0.09	OR=1.67 (0.93-3.02)	100% (<0.0001)	0.29; 9.71	No/No	Yes	NS	NA
Global cognition	5	1.9/2	590	0.11	beta=-0.04 (-0.09 to 0.01)	44.9% (0.01)	-0.13; 0.05	No/No	No	NS	NA
Time hospitalized	3	6.7/6	233	0.42	beta=-0.09 (-0.31 to 0.13)	60.3% (0.08)	-0.45; 0.27	No/No	No	NS	NA
Number of hospitalizations	3	13.1/11.1	355	0.56	beta=-0.24 (-1 to 0.57)	98.7% (<0.0001)	-1; 1	Yes/No	Yes	NS	NA
Deliberate self-harm	4	2.4/2.1	1,611	0.91	OR=1.02 (0.74-1.40)	0% (0.79)	0.74; 1.40	No/No	No	NS	NA

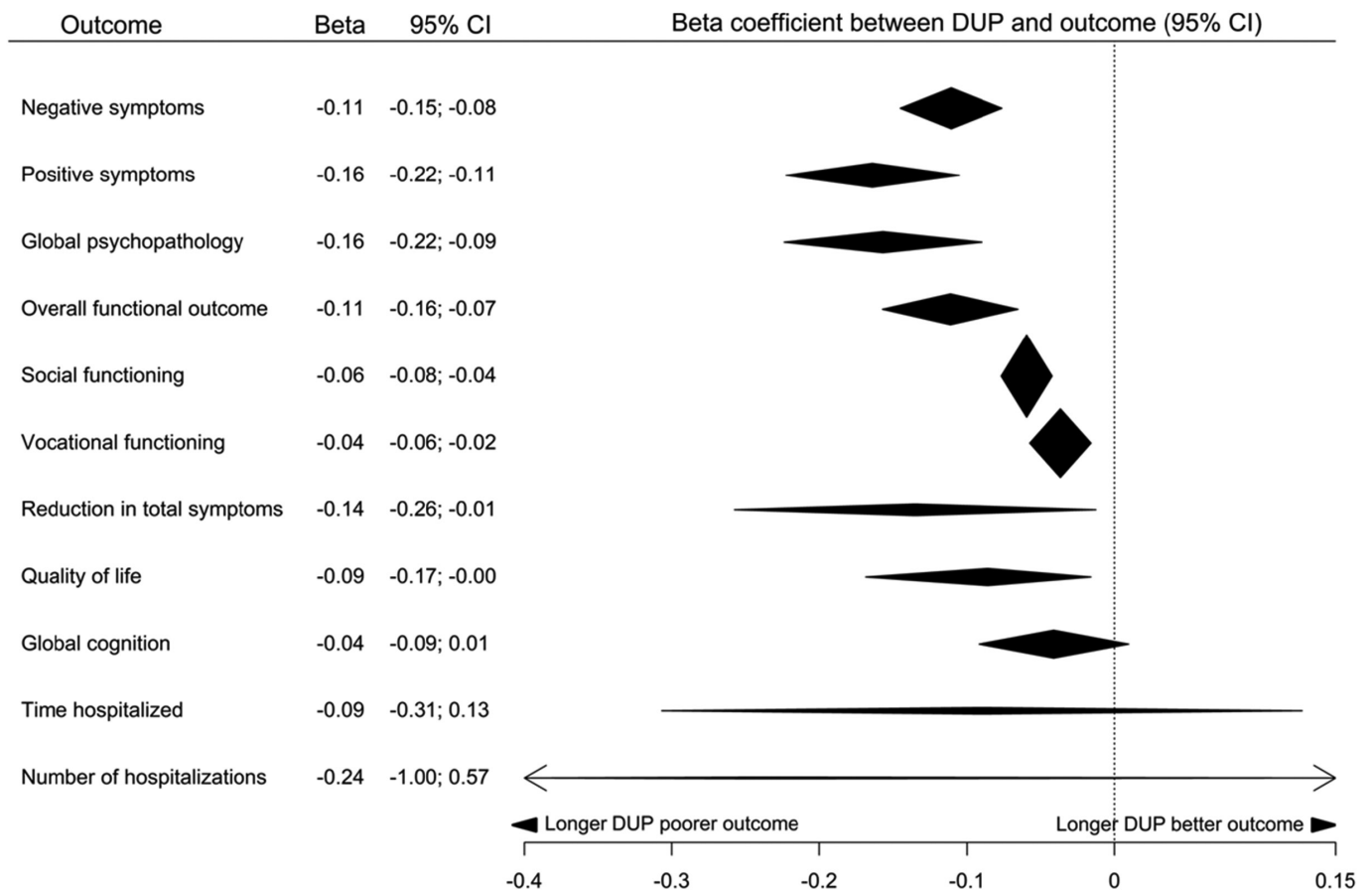OR – odds ratio, NS – not significant, NA – not applicable

Figure 5 Summary of effect sizes for relationships between duration of untreated psychosis (DUP) and continuous outcomes at follow-up

despite a more significant overall p value, due to exclusion of the largest study.

For outcomes rated class I to III, 85-95% of studies were prospective. Restricting analyses to these prospective studies led to no changes in the classes of evidence and did not significantly alter heterogeneity.

DISCUSSION

Findings and comparison with previous studies

We found highly suggestive evidence for a relationship between longer DUP and more severe positive symptoms, more severe negative symptoms and lower chance of remission at follow-up, and suggestive evidence for a relationship between longer DUP and more severe global psychopathology and poorer overall functioning at follow-up. More than 85% of studies were prospective, and these findings were all replicated in subgroup analyses restricted to prospective studies, indicating that they are unlikely to be affected by reporting bias.

There was also suggestive evidence for a relationship between longer DUP and more severe negative symptoms and higher

chance of previous self-harm at first presentation. The relationship between DUP and negative symptoms at first presentation was also evident in a subgroup analysis of antipsychotic naïve patients.

There was weak evidence for a relationship between longer DUP and poorer quality of life at first presentation and at follow-up, and also weak evidence for a relationship between longer DUP and lower chance of remission using operationalized Andreasen et al's criteria, smaller reduction in total symptoms, poorer social functioning and poorer vocational functioning at follow-up.

There was no relationship between DUP and global cognition, violence, global psychopathology, overall functioning or positive symptoms at first presentation, and between DUP and global cognition, relapse, hospitalization or deliberate self-harm at follow-up.

Our findings extend previous reviews of DUP by considering all the evidence from meta-analyses together and generating a clear hierarchy of evidence. In addition, we present the first meta-analysis of the relationship between DUP and outcomes in antipsychotic naïve patients.

Table 3 shows that each doubling in DUP predicts 8-12% more severe symptoms, and 3-8% poorer functional outcomes. Thus,
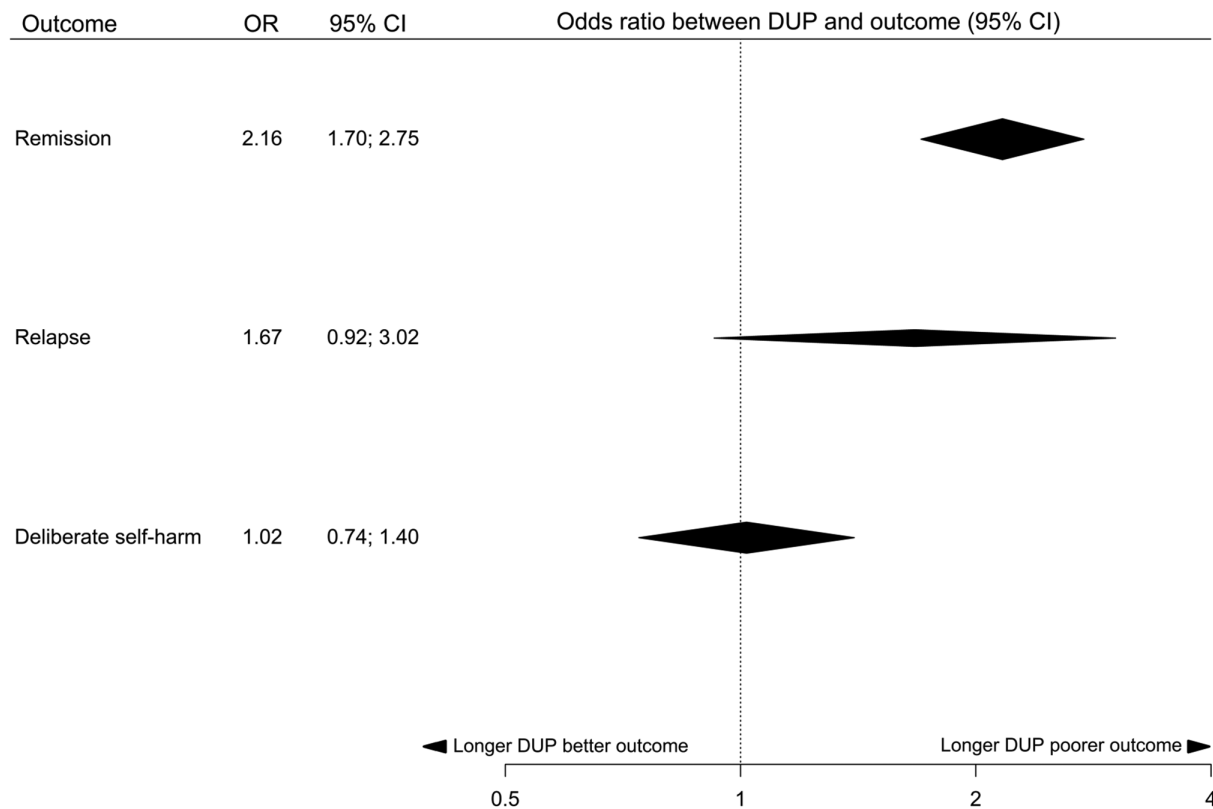
Figure 6 Summary of effect sizes for relationships between duration of untreated psychosis (DUP) and categorical outcomes at follow-up

an increase in DUP from 1 week to 4 weeks is associated with >20% more severe symptoms if the relationship is linear, which it approximates for short DUP[18,49]. This is a clinically meaningful increase. Many services have been designed worldwide with the aim of reducing DUP, and our review supports this approach by indicating that DUP is an important prognostic factor.

It is noteworthy that the largest effect size at follow-up was found between DUP and severity of positive symptoms. This suggests that the mechanism underlying positive symptoms could be central to the relationship between DUP and outcomes. Striatal dopaminergic dysfunction is thought to underlie the development of psychosis[71,72], and it has been hypothesized that psychosis feeds back on the regulation of dopamine neurons to cause further dysregulation[73,74]. Thus, a longer DUP could lead to continuing progression of dopaminergic dysfunction that makes the system less responsive to D2 antagonism when antipsychotic treatment is started[75]. However, this model does not explain more severe negative symptoms at first presentation, and we found no link between DUP and severity of positive symptoms at first presentation, which would be expected if there was a feedback loop. In addition, it remains to be determined if untreated psychosis is associated with other neurobiological changes, such as lower synaptic markers[76,77].

There are a few points of divergence from previous meta-analyses. We found a weak relationship between DUP and vocational functioning at follow-up, unlike Penttila et al[19] and

Santesteban-Echarri et al[20]. Penttila et al[19] considered a broad category of vocational functioning, which included assessments of that functioning by rating scales, real-life outcome measures (such as weeks employed or on disability pension) and binary assessments of good or poor vocational outcome based on clinician impression. Given that these assessments result in effect size measures which should not be combined in a meta-analysis, and target different underlying constructs, it is unsurprising that their results differ from our analysis. Accordingly, we encountered no significant heterogeneity in our analysis, whereas there was moderate heterogeneity in Penttila et al[19]. We defined vocational functioning as in Santesteban-Echarri et al[20]; the discrepancy with our findings is likely to be due to the inclusion, in their analysis, of a study[78] that we excluded because the sample overlapped with that of another larger included study.

Our finding of a relationship between longer DUP and more severe negative symptoms at first treatment contact is in contrast to Marshall et al[16] and Farooq et al[65], but in keeping with two larger meta-analyses[17,18]. Similarly, our finding of no relationship between DUP and first presentation positive symptoms is in contrast to the findings of Farooq et al[65], but in line with other larger meta-analyses which did not restrict inclusion criteria to low and middle income countries[16-18].

We found no relationship between DUP and risk of previous violence at first treatment contact, in contrast to the analysis by Large and Nielssen[66]. This could be explained by a unit of

analysis error in that paper, where two different outcomes which derive from the same participants (risk of violence and risk of serious violence) are combined in random-effects meta-analysis as if they were independent measures.

Strengths and limitations

Our study has several strengths, such as providing a comprehensive analysis of the relationship between DUP and clinical outcomes, and generating a clear hierarchy of evidence. We performed data extraction not just from the meta-analyses, as is common in umbrella reviews, but from the primary studies themselves, to deal with the problems of non-normally distributed data, variable reporting of different test statistics, and pooling of transformed and untransformed effect sizes, that were not addressed in many of the previous meta-analyses.

Unlike previous analyses, we used comparable outcome categories and effect sizes. Whilst the formulae used required some data imputation, which may lead to error or bias in the estimation of the effect sizes, we consider this approach preferable to exclusion of relevant studies. Sensitivity analyses indicated that our findings were robust to these data imputations, as no results went from significant to non-significant after exclusion of studies where data were imputed, and there were no significant changes in heterogeneity. Moreover, we examined the effects of DUP in antipsychotic naïve patients, and have shown for the first time that varying definitions of the endpoint in DUP moderates some of the effects observed.

We encountered considerable heterogeneity in our analyses. However, we used a random-effects model which is robust to heterogeneity[29]. The most comparable previous meta-analysis[19] also encountered moderate to substantial heterogeneity. The heterogeneity we encountered was greater, which is unsurprising as we included more studies, included studies regardless of duration of follow-up, preferred pooled results rather than schizophrenia spectrum only results if both were available, and placed no restriction on the percentage of patients with schizophrenia in our inclusion criteria.

All statistically significant results remained significant after removal of outliers. Other than remission, where the class of evidence was reduced from II to III (although with a still highly significant p value of 2×10^{-19}), all classes of evidence for significant findings remained either unchanged or were increased after removal of outliers.

Whilst our further analyses identified a number of potential contributors to heterogeneity, there remained substantial heterogeneity in first presentation quality of life, and in follow-up positive symptoms and reduction in total symptoms, which we were unable to account for. This residual heterogeneity may reflect differences in study designs, settings, outcomes and inclusion criteria.

We identified important methodological issues with previous meta-analyses. Twelve of them had critical flaws in their systematic search strategy, none were pre-registered, and only 50% performed both study selection and data extraction independently in duplicate. We attempted to mitigate these flaws as much as possible in our own meta-analysis, by pre-registering, conducting all data extraction and study selection in duplicate, and extracting all data from primary studies to ensure the fidelity of data extraction. However, as with any other meta-analysis and umbrella review, we were limited to some extent by the methodological flaws of the primary studies and meta-analyses we included.

We were reliant on the included meta-analyses to identify primary studies, and it is therefore possible that some studies were missed. However, our "file-drawer" analyses indicated that 559-1,667 null studies would be needed to negate the significant relationships we observed at both time points between DUP and negative symptoms, indicating that these findings are robust, although we also found that some other results could be sensitive to future null studies. We observed that adjusted effect sizes moderate the impact of some variables, highlighting the need to account for this aspect in future meta-analyses on DUP.

To be conservative, we categorized a sample including any treated patients as a medicated sample, as very few studies reported results separately by medication status. However, this may mean that any effect of antipsychotic treatment was diluted by the inclusion of untreated patients in some analyses. Our finding that previous antipsychotic treatment explains heterogeneity in the relationship between DUP and symptoms highlights the importance of conducting future studies at first presentation in antipsychotic naïve patients exclusively, or reporting results separately for medicated and naïve patients.

Conceptual issues in assessing DUP

We found evidence that the relationship between DUP, negative symptoms and functioning is influenced by the definition of DUP. A number of studies defined DUP as the time from the onset of psychosis to first hospitalization. Whilst this has the advantage that hospital admission is a straightforward variable, it has the disadvantage of being dependent on health service organization. However, DUP defined this way showed the strongest relationship with negative symptoms.

Another issue relating to the definition of DUP is what constitutes treatment. In some studies, it is the first dose of antipsychotic medication. However, this could be criticized, as single dose is not considered adequate treatment[79]. Some studies required 28 days of antipsychotic treatment or treatment response as the endpoint for DUP rather than initiation of treatment. Studies which used the initiation of antipsychotic treatment as the endpoint of DUP showed a stronger relationship with functional outcome than studies using adequate treatment.

These issues could be addressed through the development of operationalized criteria for DUP, as has been achieved with both remission[70] and treatment resistance[80] in psychosis.

DUP has always been assessed retrospectively in the available studies. This raises the possibility of recall bias, as patients who

are severely psychotic may have poorer long-term recall, or may attach increased significance to the transition in their mental state compared to those who are less impaired or have partially recovered. Recall bias may also be more likely as DUP becomes longer, although serial assessments of DUP during the course of clinical recovery would be needed to illuminate this aspect. Finally, recall bias may be more or less likely with different methods of ascertaining DUP, or depending on the startpoint of DUP used.

Earlier detection of psychosis may alter outcomes because the observation window is shifted (lead-time bias). Long-DUP patients may experience most of their decline in psychosocial function prior to first admission, whereas short-DUP patients may experience it after that admission[81]. It would be useful to systematically assess this potential bias in future studies.

A related issue is confounded presentation. Severe, disruptive symptoms hasten presentation and therefore shorten DUP, which could confound the relationship between DUP and variables at first presentation[49]. This may partly explain the weaker relationships in our analyses between DUP and measures at first presentation compared to follow-up measures, and could be a particular issue for our finding on deliberate self-harm. However, as longer DUP was associated with higher risk of deliberate self-harm, this confounder does not explain our finding and, if anything, would reduce the association. Nonetheless, the studies included were not well designed to address this question. Future analyses should control for severity of symptoms at first presentation to account for this potential confounder.

The studies included were all observational, which limits inferences on causation. It is possible that an unmeasured third variable explains the relationship between DUP and positive symptoms, negative symptoms, remission and functioning. Examples of potential confounding variables include premorbid adjustment and diagnosis. A meta-analysis of almost 1,400 participants found that DUP is almost four times longer in subjects with schizophrenia compared to those with affective psychosis[34]. Most studies did not report results separately for patients with affective and non-affective psychosis, but we found that diagnosis was an important moderator, with larger effect sizes for global psychopathology seen in studies with higher percentages of subjects with schizophrenia.

Moreover, it is important to consider the possibility of reverse causality. For example, our finding that a longer DUP is associated with more severe negative symptoms at first presentation could be the result of negative symptoms predating the onset of psychosis, which lead to delayed first contact with health services and persist through follow-up as they show little treatment response[77].

A further issue to take into account is that many of the outcome measures show a degree of interrelation. For example, some functional measures include assessments of symptoms, and remission is partly defined by the level of symptoms. A longitudinal modelling study showed that the effect of DUP on functional outcome measures was partly mediated by symptoms[49]. It would be useful to determine if symptom improvement mediates the relationship between DUP and other outcomes.

Adjusted effect sizes were generally smaller in the studies included in our review, which raises the possibility of selective reporting and publication of uncorrected relationships. We detected some evidence of this, with statistically significant evidence of publication bias in around 15% of our analyses. Nevertheless, no results changed from significant to non-significant and no classes of evidence changed after use of the trim-and-fill method.

It is crucial that research on DUP be designed and analyzed with confounding and reverse causality in mind. Prospective studies in people at clinical high risk, where measures can be obtained prior to the onset of the first psychotic episode, may be one approach to address these issues, albeit there will still be challenges even with such designs. For example, patients who do not engage with services, who are expected to have the longest DUP, may be unlikely to participate in these studies. Extra efforts will be required to recruit such patients and ensure representative samples.

CONCLUSIONS

The concept of DUP has contributed to a paradigm shift in psychosis services, resulting in the establishment of extensive networks of early intervention teams in many countries[11]. Our analyses show significant relationships between longer DUP and a number of important outcomes. The evidence is very suggestive for the relationships between DUP and positive symptoms, negative symptoms and chance of remission, and the effect sizes indicate that the relationships are clinically meaningful. However, more evidence is needed, particularly at first presentation and for some functional outcomes.

Future work should also investigate the mechanisms which may underlie the relationship between DUP and outcomes, explore the effect of DUP in antipsychotic naïve patients, and control for potential confounders, particularly interrelated outcome variables, mode of presentation and diagnosis, to allow clearer inferences on causation to be drawn.

REFERENCES

1. Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. World Psychiatry 2017;16:251-65.
2. McCutcheon RA, Reis Marques T, Howes OD. Schizophrenia – An overview. JAMA Psychiatry 2020;77:201-10.

3. Johnstone EC, Crow TJ, Johnson AL et al. The Northwick Park Study of first episodes of schizophrenia. I. Presentation of the illness and problems relating to admission. Br J Psychiatry 1986;148:115-20.

4. Norman RM, Malla AK. Duration of untreated psychosis: a critical examination of the concept and its importance. Psychol Med 2001;31:381-400.

5. Wyatt RJ. Neuroleptics and the natural course of schizophrenia. Schizophr Bull 1991;17:325-51.

6. Birchwood M, Todd P, Jackson C. Early intervention in psychosis. The critical period hypothesis. Br J Psychiatry 1998;172(Suppl. 33):53-9.

7. McGlashan TH. Duration of untreated psychosis in first-episode schizophrenia: marker or determinant of course? Biol Psychiatry 1999;46:899-907.

8. Killackey E, Yung AR. Effectiveness of early intervention in psychosis. Curr Opin Psychiatry 2007;20:121-5.

9. McGorry PD. Early intervention in psychosis: obvious, effective, overdue. J Nerv Ment Dis 2015;203:310-8.

10. Csillag C, Nordentoft M, Mizuno M et al. Early intervention in psychosis: from clinical intervention to health system implementation. Early Interv Psychiatry 2018;12:757-64.

11. McDaid D, Park A-L, Iemmi V et al. Growth in the use of early intervention for psychosis services: an opportunity to promote recovery amid concerns on health care sustainability. London: Personal Social Services Research Unit, 2016.

12. Omer S, Behan C, Waddington JL et al. Early intervention in psychosis: service models worldwide and the Irish experience. Irish J Psychol Med 2010;27:210-4.

13. Cheng C, Dewa CS, Langill G et al. Rural and remote early psychosis intervention services: the Gordian knot of early intervention. Early Interv Psychiatry 2014;8:396-405.

14. Verma S, Poon LY, Lee H et al. Evolution of early psychosis intervention services in Singapore. East Asian Arch Psychiatry 2012;22:114-7.

15. Mascayano F, Nossel I, Bello I et al. Understanding the implementation of coordinated specialty care for early psychosis in New York state: a guide using the RE-AIM framework. Early Interv Psychiatry 2019;13:715-9.

16. Marshall M, Lewis S, Lockwood A et al. Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. Arch Gen Psychiatry 2005;62:975-83.

17. Perkins DO, Gu H, Boteva K et al. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. Am J Psychiatry 2005;162:1785-804.

18. Boonstra N, Klaassen R, Sytema S et al. Duration of untreated psychosis and negative symptoms – a systematic review and meta-analysis of individual patient data. Schizophr Res 2012;142:12-9.

19. Penttila M, Jaaskelainen E, Hirvonen N et al. Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. Br J Psychiatry 2014;205:88-94.

20. Santesteban-Echarri O, Paino M, Rice S et al. Predictors of functional recovery in first-episode psychosis: a systematic review and meta-analysis of longitudinal studies. Clin Psychol Rev 2017;58:59-75.

21. Allott K, Fraguas D, Bartholomeusz CF et al. Duration of untreated psychosis and neurocognitive functioning in first-episode psychosis: a systematic review and meta-analysis. Psychol Med 2018;48:1592-607.

22. Bora E, Yalincetin B, Akdede BB et al. Duration of untreated psychosis and neurocognition in first-episode psychosis: a meta-analysis. Schizophr Res 2018;193:3-10.

23. Watson P, Zhang JP, Rizvi A et al. A meta-analysis of factors associated with quality of life in first episode psychosis. Schizophr Res 2018;202:26-36.

24. Fusar-Poli P, Radua J. Ten simple rules for conducting umbrella reviews. Evid Based Ment Health 2018;21:95-100.

25. Stroup DF, Berlin JA, Morton SC et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA 2000;283:2008-12.

26. Moher D, Liberati A, Tetzlaff J et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6: e1000097.

27. Harbord R. Investigating heterogeneity: subgroup analysis and meta-regression. Cochrane Statistical Methods Group Training Course, Cardiff, March 2010.

28. Petkova E, Tarpey T, Huang L et al. Interpreting meta-regression: application to recent controversies in antidepressants' efficacy. Stat Med 2013;32:2875-92.

29. Fu R, Gartlehner G, Grant M et al. Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol 2011;64:1187-97.

30. Morrison A, Polisena J, Husereau D et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care 2012;28:138-44.

31. Higgins JPT, Green S (eds). Cochrane handbook for systematic reviews of interventions. Chichester: Wiley-Blackwell, 2008.

32. Hildebrand J, Thakar S, Watts T-L et al. The impact of environmental cadmium exposure on type 2 diabetes risk: a protocol for an overview of systematic reviews. Syst Rev 2019;8:309.

33. Rohatgi A. WebPlotDigitizer. Version 4.3. Pacifica, 2017.

34. Large M, Nielssen O, Slade T et al. Measurement and reporting of the duration of untreated psychosis. Early Interv Psychiatry 2008;2:201-11.

35. Pearson K.I. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. Philos Trans R Soc Lond Series A 1900;195:1-47.

36. Dunlap WP, Burke MJ. The effect of skew on the magnitude of product-moment correlations. J Gen Psychol 1995;122:365-77.

37. Bishara AJ, Hittner JB. Confidence intervals for correlations when data are not normal. Behav Res Methods 2017;49:294-309.

38. Bonett DG. An introduction to meta-analysis. University of California, Santa Cruz, 2017.

39. Gilpin AR. Table for conversion of Kendall's tau to Spearman's rho within the context of measures of magnitude of effect for meta-analysis. Educ Psychol Meas 1993;53:87-92.

40. Rupinski MT, Dunlap WP. Approximating Pearson product-moment correlations from Kendall's tau and Spearman's rho. Educ Psychol Meas 1996; 56:419-29.

41. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.

42. Tirupati NS, Rangaswamy T, Raman P. Duration of untreated psychosis and treatment outcome in schizophrenia patients untreated for many years. Aust N Z J Psychiatry 2004;38:339-43.

43. McGorry PD, Edwards J, Mihalopoulos C et al. EPPIC: an evolving system of early detection and optimal management. Schizophr Bull 1996;22:305-26.

44. Jacobs P, Viechtbauer W. Estimation of the biserial correlation and its sampling variance for use in meta-analysis. Res Synth Methods 2017;8:161-80.

45. Borenstein M, Hedges LV, Higgins JPT et al (eds). Introduction to meta-analysis. Chichester: Wiley, 2009.

46. Bonett DG. Transforming odds ratios into correlations for meta-analytic research. Am Psychol 2007;62:254-5.

47. Cheng Y, Liu H. A short note on the maximal point-biserial correlation under non-normality. Br J Math Stat Psychol 2016;69:344-51.

48. Souverein OW, Dullemeijer C, van't Veer P et al. Transformations of summary statistics as input in meta-analysis for linear dose-response models on a logarithmic scale: a methodology developed within EURRECA. BMC Med Res Methodol 2012;12:57.

49. Drake RJ, Husain N, Marshall M et al. Effect of delaying treatment of first-episode psychosis on symptoms and social outcomes: a longitudinal analysis and modelling study. Lancet Psychiatry 2020;7:602-10.

50. Ristić-Medić D, Dullemeijer C, Tepsić J et al. Systematic review using meta-analyses to estimate dose-response relationships between iodine intake and biomarkers of iodine status in different population groups. Nutr Rev 2014;72:143-61.

51. Moran VH, Stammers AL, Medina MW et al. The relationship between zinc intake and serum/plasma zinc concentration in children: a systematic review and dose-response meta-analysis. Nutrients 2012;4:841-58.

52. Tenny S, Hoffman MR. Odds ratio (OR). StatPearls Publishing, 2020.

53. Shor E, Roelfs D, Vang ZM. The "Hispanic mortality paradox" revisited: meta-analysis and meta-regression of life-course differentials in Latin American and Caribbean immigrants' mortality. Soc Sci Med 2017;186:20-33.

54. Viechtbauer W. Conducting meta-analyses in R with the metafor Package. J Stat Softw 2010;36.

55. Egger M, Smith GD, Schneider M et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629.

56. Radua J, Ramella-Cravaro V, Ioannidis JPA et al. What causes psychosis? An umbrella review of risk and protective factors. World Psychiatry 2018;17:49-66.

57. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000;56:455-63.

58. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. Clin Trials 2007;4:245-53.

59. Kossmeier M, Tran US, Voracek M. metaviz. Version 0.3.1. https://cran.r-project.org/.

60. Rosenberg MS. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. Evolution 2005;59:464-8.

61. Harrer M, Cuijpers P, Furukawa TA et al. Doing meta-analysis in R: a hands-on guide. PROTECT Lab Erlangen, 2019.

62. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289-300.

63. Kahn RS, Winter van Rossum I, Leucht S et al. Amisulpride and olanzapine followed by open-label treatment with clozapine in first-episode schizophrenia and schizophreniform disorder (OPTiMiSE): a three-phase switching study. Lancet Psychiatry 2018;5:797-807.

64. Oliver D, Davies C, Crossland G et al. Can we reduce the duration of untreated psychosis? A systematic review and meta-analysis of controlled interventional studies. Schizophr Bull 2018;44:1362-72.

65. Farooq S, Large M, Nielssen O et al. The relationship between the duration of untreated psychosis and outcome in low-and-middle income countries: a systematic review and meta analysis. Schizophr Res 2009;109:15-23.

66. Large MM, Nielssen O. Violence in first-episode psychosis: a systematic review and meta-analysis. Schizophr Res 2011;125:209-20.

67. Alvarez-Jimenez M, Priede A, Hetrick SE et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012;139:116-28.

68. Burns JK. Cannabis use and duration of untreated psychosis: a systematic review and meta-analysis. Curr Pharm Des 2012;18:5093-04.

69. Challis S, Nielssen O, Harris A et al. Systematic meta-analysis of the risk factors for deliberate self-harm before and after treatment for first-episode psychosis. Acta Psychiatr Scand 2013;127:442-54.

70. Andreasen NC, Carpenter WT Jr, Kane JM et al. Remission in schizophrenia: proposed criteria and rationale for consensus. Am J Psychiatry 2005;162:441-9.

71. Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III – the final common pathway. Schizophr Bull 2009;35:549-62.

72. Brugger SP, Angelescu I, Abi-Dargham A et al. Heterogeneity of striatal dopamine function in schizophrenia: meta-analysis of variance. Biol Psychiatry 2020;87:215-24.

73. McCutcheon RA, Krystal JH, Howes OD. Dopamine and glutamate in schizophrenia: biology, symptoms and treatment. World Psychiatry 2020;19:15-33.

74. Howes OD, Murray RM. Schizophrenia: an integrated sociodevelopmental-cognitive model. Lancet 2014;383:1677-87.

75. Potkin SG, Kane JM, Correll CU et al. The neurobiology of treatment-resistant schizophrenia: paths to antipsychotic resistance and a roadmap for future research. NPJ Schizophr 2020;6:1.

76. Onwordi EC, Halff EF, Whitehurst T et al. Synaptic density marker SV2A is reduced in schizophrenia patients and unaffected by antipsychotics in rats. Nat Commun 2020;11:246.

77. Osimo EF, Beck K, Reis Marques T et al. Synaptic loss in schizophrenia: a meta-analysis and systematic review of synaptic protein and mRNA measures. Mol Psychiatry 2019;24:549-61.

78. Tandberg M, Ueland T, Sundet K et al. Neurocognition and occupational functioning in patients with first-episode psychosis: a 2-year follow-up study. Psychiatry Res 2011;188:334-42.

79. Kaar SJ, Natesan S, McCutcheon R et al. Antipsychotics: mechanisms underlying clinical response and side-effects and novel treatment approaches based on pathophysiology. Neuropharmacology 2020;172:107704.

80. Howes OD, McCutcheon R, Agid O et al. Treatment-resistant schizophrenia: Treatment Response and Resistance in Psychosis (TRRIP) Working Group Consensus Guidelines on Diagnosis and Terminology. Am J Psychiatry 2017;174:216-29.

81. Jonas KG, Fochtmann LJ, Perlman G et al. Lead-time bias confounds association between duration of untreated psychosis and illness course in schizophrenia. Am J Psychiatry 2020;177:327-34.