A systematic review of longitudinal outcome studies of first-episode psychosis

N. M. MENEZES*, T. ARENOVICH AND R. B. ZIPURSKY

Schizophrenia Program, Centre for Addiction and Mental Health, Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Background. Existing outcome literature has had an over-representation of chronic patients and suggested a progressive course and poor outcome for schizophrenia. The current study aimed to recombine data of samples from longitudinal studies of first-episode psychosis (FEP) to describe outcome and its predictors.

Method. A literature search (1966–2003) was conducted for prospective studies examining outcome in first-episode non-affective psychosis using the following key words: early, first, incident, episode, admission, contact, psychosis, schizophrenia, psychotic disorders, course, outcome, follow-up, longitudinal, cohort. These were pooled and analyzed using descriptive and regression analyses.

Results. Thirty-seven studies met the inclusion criteria, representing 4100 patients with a mean follow-up of 35.1 ± 6.0 months. Studies varied in the categories of outcome used, the most common being ‘good’ (54% of studies) and ‘poor’ (34% of studies), variably defined. In studies reporting these categories, good outcomes were reported in 42.2% (3.5%) and poor outcomes in 27.1% (2.8%) of cases. Predictors associated with better outcome domains were: combination of pharmacotherapy and psychosocial therapy, lack of epidemiologic representativeness of the sample, and a developing country of origin. Use of typical neuroleptics was associated with worse outcome. Stratification analyses suggested that populations with schizophrenia only, and those with prospective design, were associated with worse outcome domains.

Conclusions. Outcome from FEP may be more favorable than previously reported, and treatment and methodological variables may be important contributors to outcome. Significant heterogeneity in definitions and methodology limited the comparison and pooling of data. A multi-dimensional, globally used definition of outcome is required for future research.

INTRODUCTION

Research into the outcome from schizophrenia has demonstrated variability in definitions of sample, outcomes and measures, making comparison difficult (Wing, 1988; Liberman et al. 2002). A heterogeneous outcome from schizophrenia has been reported, with a smaller proportion (20–50%) experiencing recovery or significant improvement, compared to a majority with a course of multiple episodes and increasing impairment (Bleuler, 1978; Ciompi, 1980; Huber et al. 1980; Shepherd et al. 1989; Harding et al. 1992; van Os et al. 1996). This was confirmed in a meta-analysis of the schizophrenia outcome literature, which reported that only 40% of patients were improved after
follow-up averaging 5.6 years (Hegarty et al. 1994).

The existing literature on schizophrenia has been influenced by issues related to sample representativeness. The prevalent samples have included patients at different stages of illness, with an over-representation of chronic, treatment-refractory patients (Shepherd et al. 1989; Keshavan & Schooler, 1992; Lieberman et al. 1996; Birchwood et al. 1998; Riecher-Rossler & Rossler, 1998), and an attrition of patients who have done well or recovered but would have met an initial diagnosis of schizophrenia (Davidson & McGlashan, 1997). These factors bias results toward a poor outcome and have influenced current clinical and public perspectives on the prognosis of schizophrenia. This exemplifies the ‘clinician’s illusion’, involving the attribution of the characteristics and course of those patients who are currently ill to the entire population contracting the illness (Cohen & Cohen, 1984).

The current systematic review aims to address sampling issues in the existing literature by analyzing first-episode psychosis (FEP) outcome studies. Given patients are at similar stages of illness, the outcomes from such studies may be considered more generalizable than those from more chronic samples. The studies are combined and reanalyzed in the style of a meta-analysis, with the goal of examining outcome and predictors in a large patient sample. The study hypothesizes that analysis of incident samples of psychosis will demonstrate a more favorable outcome from schizophrenia-spectrum disorders than previously reported.

**METHOD**

**Data sources and selection**

Relevant studies were identified by searching MEDLINE, PSYCHINFO, and the Cochrane Database of Systematic Reviews, 1966 to December 2003. The key words used, in various combinations, were: early, first, incident, episode, admission, contact, psychosis, schizophrenia, psychotic disorders, course, outcome, follow-up, longitudinal, prospective, cohort. Further studies were obtained by manual reference examination of published reports and citations of unpublished research. Non-English studies were evaluated through translators. The inclusion criteria applied to studies were:

1. **FEP**, defined as patients who are: making their first treatment contact for psychotic symptoms or in their first admission for psychotic symptoms or in their first episode of psychosis and in the absence of an affective disorder (i.e. only schizophrenia-spectrum diagnoses included).
2. The standardized diagnostic system in use must be specified (e.g. RDC, DSM).
3. Study criteria include minimum age 14 years (to maximize the number of patients without including childhood-onset cases).
4. A prospective follow-up of at least 6 months.
5. Follow-up must yield adequate outcome data for analysis (in the clinical/functional/personal domain); e.g. studies reporting outcomes exclusively in the domains of cognition, neuroimaging, suicide or treatment adherence, or studies that reported no raw data amenable to reanalysis (i.e. only reported correlations, or significance of difference between groups), would not meet this criterion.

The exclusion criteria for studies were:

1. Study sample included organic etiology of psychosis.
2. Sample included substance-induced psychosis, no separate data provided.
3. Sample included mental retardation.
4. Diagnosis made retrospectively based on chart review.
5. Categorical outcome (e.g. remission, relapse or response) not explicitly defined.

Our search (Fig. 1) yielded 37 study cohorts for final analysis (Table 1). The studies meeting inclusion criteria were all published after 1980, with the highest number of studies published in the year 2000 and then a decreasing number per annum since that time.

In keeping with the study’s goal of combining results from as many samples as possible, a minimum sample size criterion was not used, although studies were weighted by sample size. Multiple studies on the same patient cohort were accepted if there were different outcome
measures or time points contributing to separate analyses. Individual sites that had multiple reports with overlap of patients were coded so that the maximum amount of data from the cohort was used without duplication of patients. If this could not be done, individual reports were excluded. Multi-site studies were included, with data reported for the whole cohort and not for the individual sites.

Studies including patients with affective psychosis were included if there was follow-up confirmation of diagnosis and separate data were provided for the non-affective group. Similarly, studies with multiple diagnoses at follow-up were included only if separate outcome data for schizophrenia-spectrum disorders were provided. Studies including non-FEP patients were included if separate data were provided for the FEP group.

Data analysis
An extraction form was created following a review of the literature on different variables contributing to outcome. Studies were reviewed and the following variables were extracted and included in univariate analyses: decade of publication, country of origin, epidemiological representativeness of the sample (e.g. representative samples came from a catchment area, and were not selected from a biased source such as a private hospital), design, study follow-up duration, patient status (in- or out-patient), gender, cohort treatment-naive at study entry, prodrome duration, duration of untreated psychosis, substance-use exclusion criteria, type of therapy (pharmacotherapy, psychosocial therapy, combination, or not specified), age of onset, age at study entry, and percentage compliance.

Studies reported on different outcome measures (rates in Table 2); one unifying outcome for analysis was not possible. Outcome variables used in <10% of studies were not analyzed. Descriptive analyses were used for outcomes, listed in Table 2. Many studies used categorical outcome measures based, for example, on symptom scale scores or numbers of admissions (e.g. good/intermediate/poor, remitted/improved/chronic). Thus, most outcome data were originally reported in 1–3 categories in various combinations of good/intermediate/poor.

A series of preliminary graphs and univariate analyses [weighted repeated measures analyses of variance (ANOVAs) for categorical predictors, weighted repeated measures linear regression analyses for continuous predictors] was performed to determine predictor variable inclusion and contribution to variability (all univariate analyses were carried out as a function of outcome, and thus were technically bivariate). Studies were weighted according to the follow-up sample size of each cohort and the retention rate, at the corresponding time point. Factors found to be associated with the outcome measures in the univariate analyses (threshold of \( p \leq 0.10 \) for inclusion) were then combined and tested in multiple regression (through a series of repeated measures linear models). To examine the effect of contributing variables of interest, including potential confounders, the analyses were then stratified by the following variables: epidemiological representativeness (yes/no), follow-up duration (\( \leq \)/\( > \)2 years), diagnosis (schizophrenia/other), design [randomized controlled trial (RCT) and quasi-experimental/other]. Using more than two levels of stratification was not possible due to power limitations.
Table 1. Summary of included studies

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N*</th>
<th>Drop-out rate (%)</th>
<th>F/U</th>
<th>Diagnosis</th>
<th>Outcome definition</th>
<th>Study design</th>
<th>Independent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kane et al. (1982)</td>
<td>19</td>
<td>35 (whole cohort)</td>
<td>12</td>
<td>FES (subgroup)</td>
<td>Relapse (= substantial clinical deterioration)</td>
<td>RCT</td>
<td>Fluphenazine versus placebo</td>
</tr>
<tr>
<td>Crow et al. (1986); Geddes et al. (1994); Johnstone et al. (1986, 1990)</td>
<td>120</td>
<td>10-8</td>
<td>24-175</td>
<td>FEP (subgroup Scz for RCT)</td>
<td>Relapse = readmission, employment, total hospitalization time</td>
<td>Prospective (RCT subset)</td>
<td>Treatment versus placebo, DUP, social and behavioral measures, neurological measures, demographics, depression, symptoms</td>
</tr>
<tr>
<td>Rabine et al. (1986)</td>
<td>36</td>
<td>22-2</td>
<td>12</td>
<td>FES (Scz subgroup)</td>
<td>Relapse, remission (no sx for 3 months), three outcome categories: remission, relapsed, in episode (no remission)</td>
<td>Prospective</td>
<td>Medication, pre-morbid functioning, illness duration</td>
</tr>
<tr>
<td>Scottish Schizophrenia Research Group (1987, 1988, 1992); McCreadie et al. (1989)</td>
<td>49</td>
<td>20-4-26-5</td>
<td>12, 24, 60</td>
<td>FES</td>
<td>Outcome good = no relapses/sx, poor = relapse and/or sx at follow-up; unemployment, remission, readmission</td>
<td>Prospective</td>
<td>Gender, symptoms, diagnosis, placebo, employment</td>
</tr>
<tr>
<td>Lieberman et al. (1989, 1992, 1993); Loebel et al. (1992); Robinson et al. (1999); Szymanski et al. (1995)</td>
<td>54–104</td>
<td>1-8–11-9</td>
<td>12–60</td>
<td>FES (subset of a larger sample)</td>
<td>Improved/remitted, course, treatment response based on CGI, SADS, SANS (full, partial, none, stabilized)</td>
<td>Prospective</td>
<td>Gender, treatment, symptoms, social adjustment, demographics, treatment</td>
</tr>
<tr>
<td>Rajkumar &amp; Thara (1989)</td>
<td>96</td>
<td>22-9</td>
<td>36</td>
<td>FES</td>
<td>Relapse (re-emergence or worsening of sx), re-hospitalization; three outcomes: non-relapsing, relapsing, continuously ill</td>
<td>Prospective</td>
<td>Mood symptoms, social functioning, diagnosis, compliance</td>
</tr>
<tr>
<td>Shepherd et al. (1989)</td>
<td>49</td>
<td>11-6 (whole cohort)</td>
<td>60</td>
<td>FES (male and female subgroups)</td>
<td>Employment, symptoms, course, readmission, social functioning, mortality, remission = one episode + no impairment, improved = several episodes and no/minimum impairment, poor = no return to normality</td>
<td>Prospective</td>
<td>Demographics (particularly gender)</td>
</tr>
<tr>
<td>Barrelet et al. (1990)</td>
<td>51</td>
<td>9-8</td>
<td>9</td>
<td>FES</td>
<td>Relapse (recurrence or exacerbation of symptoms at least 1 month after discharge)</td>
<td>Prospective</td>
<td>Expressed emotion, demographics</td>
</tr>
<tr>
<td>Helgason (1990)</td>
<td>107</td>
<td>1-9</td>
<td>240</td>
<td>FES</td>
<td>Employment, readmission, social functioning, symptoms (none or minimal, obvious, severe)</td>
<td>Prospective</td>
<td>Demographics, social functioning, treatment</td>
</tr>
<tr>
<td>Salokangas &amp; Stengard (1990)</td>
<td>227</td>
<td>4-3–12-6</td>
<td>24</td>
<td>FEP</td>
<td>Symptoms, employment, social relationships, GAS, sexual development</td>
<td>Prospective</td>
<td>Demographics, pre-morbid development, autonomy</td>
</tr>
<tr>
<td>Delisi et al. (1992)</td>
<td>30</td>
<td>3-3</td>
<td>24</td>
<td>FEP</td>
<td>Hospitalization amount, GAS, BPRS, SCS</td>
<td>Prospective</td>
<td>Brain changes</td>
</tr>
<tr>
<td>Jablensky et al. (1992)</td>
<td>687</td>
<td>21-8 (whole cohort)</td>
<td>24</td>
<td>FEP (Scz subgroup)</td>
<td>Remission, remission with residual symptoms, unremitting</td>
<td>Prospective</td>
<td>Developed versus developing countries</td>
</tr>
<tr>
<td>Tohen et al. (1992, 2000); Zarate et al. (2000)</td>
<td>85</td>
<td>8-27</td>
<td>6, 24</td>
<td>FEP (subset of non-affective disorders)</td>
<td>Recovery (= BPRS cut-offs for &gt;8 weeks, CGI &lt;2; and functional = GAF, etc.), recurrence as differentiated from relapse</td>
<td>Prospective</td>
<td>Diagnosis, demographics, onset type, symptoms, co-morbidity</td>
</tr>
<tr>
<td>Flaum et al. (1992); Ho et al. (1998, 2000)</td>
<td>74</td>
<td>?</td>
<td>6</td>
<td>FEP (subset)</td>
<td>Symptom severity (SANS, SAPS), symptom remission, poor outcome (marked impairment in social adjustment and GAS &lt;40) quality of life (employment, etc.)</td>
<td>Prospective</td>
<td>DUP, demographics, symptom dimensions</td>
</tr>
<tr>
<td>Thara et al. (1994)</td>
<td>90</td>
<td>15-6</td>
<td>120</td>
<td>FES</td>
<td>Course (put into three categories, based on numbers of relapse = reappearance of ≥ 1 sx after 1 month remission, remission = absence of psychosis sx), symptoms, suicide, living status</td>
<td>Prospective</td>
<td>Demographics</td>
</tr>
<tr>
<td>Zhang et al. (1994)</td>
<td>83</td>
<td>6-0</td>
<td>18</td>
<td>FES (all male)</td>
<td>Readmission, hospital-free period, BPRS, GAF</td>
<td>RCT</td>
<td>Family intervention versus standard care, compliance</td>
</tr>
<tr>
<td>Huguelet et al. (1995)</td>
<td>67</td>
<td>41-7</td>
<td>48</td>
<td>FES</td>
<td>GAF (good/bad, cut-off 51), course (good and stable, fluctuation, bad and stable)</td>
<td>Prospective</td>
<td>Demographics, relatives’ EE, disability</td>
</tr>
<tr>
<td>Zhang-Wong et al. (1995)</td>
<td>7-49</td>
<td>?</td>
<td>60</td>
<td>FEP (FES and scF subgroups only)</td>
<td>Employment, living status, treatment compliance, GAF, readmission</td>
<td>Prospective</td>
<td>Diagnosis</td>
</tr>
<tr>
<td>Bremet et al. (1996); Craig et al. (2000)</td>
<td>96-219</td>
<td>10-29-2</td>
<td>6, 24</td>
<td>FEP (FES subgroup only)</td>
<td>Rehospitalization, employment, remission (full/partial/not), three categories course: episodes + complete remission, episodes + partial remission, continuously ill, functioning (GAF)</td>
<td>Prospective</td>
<td>Diagnosis, pre-morbid adjustment, DUP, symptoms</td>
</tr>
<tr>
<td>Authors</td>
<td>Sample Size</td>
<td>F/U (months)</td>
<td>Study Design</td>
<td>Outcome</td>
<td>Independent Variables</td>
<td>Design</td>
<td>Demographics</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------</td>
<td>-----------------------</td>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>Wieselgren &amp; Lindstrom (1996)</td>
<td>66</td>
<td>4.5–24.2</td>
<td>Prospective</td>
<td>Good/poor/intermediate outcome (SCS), suicide, readmission, employment</td>
<td>FES</td>
<td>Demographics, pre-morbid functioning</td>
<td></td>
</tr>
<tr>
<td>DeLisi et al. (1998)</td>
<td>50</td>
<td>0–60</td>
<td>Prospective</td>
<td>Scores (based on BPRS, GAS, functioning) into three outcomes (complete, partial and no recovery); brain volume changes</td>
<td>FES</td>
<td>Neuroimaging, pre-morbid history, demographics</td>
<td></td>
</tr>
<tr>
<td>Edwards et al. (1998)</td>
<td>198</td>
<td>7–5</td>
<td>Prospective</td>
<td>Prolonged recovery = failure to sustain remission (minimum one BPRS item at time 2, 3, 4)</td>
<td>FEP</td>
<td>Demographics, DUP, depression, substance use, psychosocial functioning</td>
<td></td>
</tr>
<tr>
<td>Gur et al. (1998)</td>
<td>20</td>
<td>0–30 (mean)</td>
<td>Prospective</td>
<td>Brain volumes, symptom severity (SANS, SAPS)</td>
<td>FES</td>
<td>Brain imaging and neurobehavioral studies, medications</td>
<td></td>
</tr>
<tr>
<td>Takei et al. (1998)</td>
<td>88</td>
<td>7–6</td>
<td>Prospective</td>
<td>Diagnosis, symptoms, readmission, duration hospitalization, SAS, GAS</td>
<td>FEP</td>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Larsen et al. (2000)</td>
<td>43</td>
<td>0–12</td>
<td>Prospective</td>
<td>Remission (PANSS &lt; 2 for &gt; 2 months), GAF, three outcomes (remitted, relapse type, continuously psychotic)</td>
<td>FEP</td>
<td>DUP, pre-morbid functioning, gender</td>
<td></td>
</tr>
<tr>
<td>Lehtinen et al. (2000)</td>
<td>135</td>
<td>21–8</td>
<td>Prospective</td>
<td>Time in hospital, symptoms (BPRS), remission (no psychotic sx for 1 year), employment, GAF, GOL</td>
<td>FEP</td>
<td>Minimal versus usual practice neuroleptics</td>
<td></td>
</tr>
<tr>
<td>de Lint et al. (2000)</td>
<td>53</td>
<td>39–6</td>
<td>Prospective</td>
<td>Symptoms (BPRS improved by 5 points)</td>
<td>FES</td>
<td>Cognitive deficits</td>
<td></td>
</tr>
<tr>
<td>Singh et al. (2000)</td>
<td>56</td>
<td>0–36</td>
<td>Prospective</td>
<td>Rehospitalization, employment, GAF, remission, Bleuler’s outcomes, course relapse (BPRS), social functioning; good = no relapse, intermediate = more than one relapse, poor = chronic positive symptoms</td>
<td>FEP (subgroup of FES)</td>
<td>Diagnosis (affective versus substance versus non-affective)</td>
<td></td>
</tr>
<tr>
<td>Linszen et al. (2001)</td>
<td>76</td>
<td>3–9</td>
<td>Prospective</td>
<td>CAN, number of hospital days</td>
<td>FEP</td>
<td>Early intervention</td>
<td></td>
</tr>
<tr>
<td>Cahn et al. (2002)</td>
<td>34</td>
<td>14–7</td>
<td>Prospective</td>
<td>Drop-out, relapse (change in BPRS ≥ 10, CGI ≥ 6, GAS ≥ 20), rehospitalization, social, compliance, side-effects</td>
<td>FES</td>
<td>Brain volume changes, symptoms, DUP, antipsychotic use</td>
<td></td>
</tr>
<tr>
<td>Gaebel et al. (2002)</td>
<td>115</td>
<td>56–5</td>
<td>RCT</td>
<td>Drop-out, relapse (change in BPRS ≥ 10, CGI ≥ 6, GAS ≥ 20), rehospitalization, social, compliance, side-effects</td>
<td>FES</td>
<td>Intermittent versus continuous treatment</td>
<td></td>
</tr>
<tr>
<td>Moller et al. (2002)</td>
<td>291</td>
<td>61–6</td>
<td>Prospective</td>
<td>Negative syndrome (SANS), readmission, duration hospitalization, severe sx and functioning (GAF &lt; 51), severity (CGI, PANSS)</td>
<td>FEP</td>
<td>Negative syndrome, diagnosis (affective versus non-affective)</td>
<td></td>
</tr>
<tr>
<td>Novak-Grubic &amp; Tavcar (2002)</td>
<td>56</td>
<td>0–12</td>
<td>Prospective</td>
<td>Non-compliance, relapse (clinical judgment and increased medications)</td>
<td>FEP (male only)</td>
<td>Symptoms, side-effects, demographics, diagnosis, insight</td>
<td></td>
</tr>
<tr>
<td>Whitehorn et al. (2002)</td>
<td>103</td>
<td>45–6–52.4</td>
<td>Prospective</td>
<td>Recovery (symptomatic = PANSS items score &lt; 3, functional = SOFAS &gt; 60, GAF &gt; 50)</td>
<td>FEP (compliant only)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lieberman et al. (2003)</td>
<td>160</td>
<td>37–5</td>
<td>Prospective</td>
<td>Time to/in remission (50 % decrease in BPRS, CGI ≤ 3)</td>
<td>FES/sczP</td>
<td>Clorapine versus chlorpromazine, DUP</td>
<td></td>
</tr>
<tr>
<td>Stirling et al. (2003)</td>
<td>62</td>
<td>18–8</td>
<td>RCT</td>
<td>Symptoms (SANS, SAPS), GAF, employment, readmission</td>
<td>FEP</td>
<td>Neurocognitive functioning, insight, symptoms</td>
<td></td>
</tr>
<tr>
<td>Addington et al. (2003a, b; 2004); Coldham et al. (2002)</td>
<td>253–290</td>
<td>17–9–30</td>
<td>Prospective</td>
<td>Suicidal behavior, PANSS, CDS, remission (PANSS ≤ 3 positive items), social functioning (QLS), medication adherence</td>
<td>FEP (non-affective)</td>
<td>Demographics, depression, symptoms, social functioning, substance use, quality of life, side-effects, family involvement</td>
<td></td>
</tr>
</tbody>
</table>

*a Sample size at entry, x1.
*b Follow-up (duration) in months.
*c Diagnoses abbreviations: FEP, first-episode psychosis (any combination of schizophrenia, schizophreniform, schizo-affective, delusional or NOS disorder; FES, first-episode schizophrenia; sczP, schizophreniform disorder.
*d Outcome abbreviations: BPRS, Brief Psychiatric Rating Scale; CAN, Camberwell Assessment of Need; CGI, Clinical Global Impression Scale; GOL, Grip on Life Questionnaire; PANSS, Positive and Negative Syndrome Scale; CDS, Calgary Depression Scale for Schizophrenia; QLS, Quality of Life Scale; GAF/GAS, Global Assessment of Functioning Scale; SADS, Schedule for Affective Disorders and Schizophrenia; SAS, Social Adjustment Scale; SOFAS, Social and Occupational Functioning Assessment Scale; SCS, Strauss–Carpenter Scale; sx, symptoms.
*e Study design abbreviations: RCT, randomized controlled trial. 
*f Independent variables abbreviations: DUP, duration of untreated psychosis; EE, expressed emotion.
All statistical analyses were carried out using the SAS System v. 8.2. (SAS Institute Inc., Cary, NC, USA). The level of significance for all analyses (except univariate) was set at $p \leq 0.05$.

RESULTs

Study selection and demographics

Our search identified 37 cohorts representing a sample size of approximately 4100 patients (varying according to time point and variable). Reasons for study exclusion are listed in Fig. 1. The mean follow-up duration was $35.1 \pm 6.0$ months. Fifty-eight per cent of studies had follow-up of 6 months to 2 years with a mean retention rate of 86%, while 42% of studies followed patients for more than 2 years with a mean retention rate of 80%. Eighty-five per cent of the included studies originated in developed countries. Fifty-four per cent of the included studies reported data exclusively on patients with a diagnosis of schizophrenia, with an additional 20% reporting on schizophrenia-spectrum disorders (e.g. schizophreniform and schizo-affective disorders).

The mean age at study entry was $27.3 \pm 3.9$ years and at illness onset was $25.7 \pm 4.3$ years. Males represented $63.0 \pm 21.0\%$ of the sample. The mean duration of untreated psychosis was $18.4 \pm 10.7$ months.

Outcome

Differentially defined outcomes were reported across studies (Table 2), with the most common being ‘good’ and ‘poor’ outcomes (grossly defined per study in Table 1). Studies reported these categories singularly (‘good outcome’ in 25% of studies, ‘intermediate’ in 1%, and ‘poor’ in 6%) and in combination (‘good + poor’ in 4% of studies and ‘good + intermediate + poor in 24%). For those studies reporting these categories, good outcome was described in 42% of the population, and poor outcome in 27%. The heterogeneity of predictor/outcome definitions reduced the comparability and sample size for each variable, limiting the number of possible analyses and the validity and relevance of conclusions, despite a large combined sample size.

Predictors of outcome

Outcome measures not significantly associated with any independent variables in univariate analysis (functional recovery, intermediate outcome, relapse) were removed from further analyses. The following outcomes were used for further multivariable models (for the studies...
that reported them): good, poor, readmission, Global Assessment of Functioning Scale (GAF), employment/education (Table 3). The estimates of various outcomes are summarized by predictor variables in Table 4.

**Table 3. Multivariable model results**

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Independent variable</th>
<th>$F$ statistic</th>
<th>Degrees of freedom</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good outcome</td>
<td>Age</td>
<td>0.33</td>
<td>1, 7</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Combination therapy</td>
<td>4.29</td>
<td>1, 18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Design</td>
<td>&lt;0.01</td>
<td>1, 18</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>EPI representative</td>
<td>8.26</td>
<td>1, 18</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Follow-up duration</td>
<td>5.25</td>
<td>1, 8</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Origin type</td>
<td>12.66</td>
<td>1, 18</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Poor outcome</td>
<td>Pharmacotherapy</td>
<td>16.68</td>
<td>2, 14</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Tx-naive at entry</td>
<td>10.78</td>
<td>2, 14</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>GAF</td>
<td>3.77</td>
<td>1, 9</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Study design</td>
<td>2.75</td>
<td>1, 9</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>EPI representative</td>
<td>0.34</td>
<td>1, 9</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>Origin type</td>
<td>0.54</td>
<td>1, 9</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Substance use permitted</td>
<td>0.12</td>
<td>2, 9</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Tx at entry</td>
<td>21.70</td>
<td>1, 9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Combination therapy</td>
<td>6.55</td>
<td>1, 8</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>EPI representative</td>
<td>9.81</td>
<td>1, 8</td>
<td>0.01</td>
</tr>
</tbody>
</table>

GAF, Global Assessment of Functioning Scale; EPI, epidemiological; Tx, treatment.

**Table 4. Summary of estimates of outcomes by significant predictor variables**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Predictor variable</th>
<th>% Outcome ($\pm$ s.e.)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good outcome</td>
<td>Combination therapy – yes</td>
<td>54.7 ± 8.3</td>
<td>38.4–71.0</td>
</tr>
<tr>
<td></td>
<td>Combination therapy – N.S.</td>
<td>37.6 ± 3.5</td>
<td>30.7–44.5</td>
</tr>
<tr>
<td></td>
<td>EPI representative – yes</td>
<td>35.6 ± 5.6</td>
<td>24.6–46.6</td>
</tr>
<tr>
<td></td>
<td>EPI representative – no</td>
<td>47.1 ± 4.4</td>
<td>38.5–55.7</td>
</tr>
<tr>
<td></td>
<td>Country of origin – developed</td>
<td>38.8 ± 3.6</td>
<td>31.7–45.9</td>
</tr>
<tr>
<td></td>
<td>Country of origin – developing</td>
<td>48.1 ± 11.5</td>
<td>25.6–70.6</td>
</tr>
<tr>
<td>Poor outcome</td>
<td>Pharmacotherapy – typicals</td>
<td>53.0 ± 7.0</td>
<td>39.3–66.7</td>
</tr>
<tr>
<td></td>
<td>Pharmacotherapy – atypicals</td>
<td>26.4 ± 8.8</td>
<td>9.2–43.6</td>
</tr>
<tr>
<td></td>
<td>Pharmacotherapy – N.S.</td>
<td>25.6 ± 2.8</td>
<td>20.1–31.1</td>
</tr>
<tr>
<td></td>
<td>Tx-naive at entry – yes</td>
<td>47.3 ± 13.2</td>
<td>21.4–73.2</td>
</tr>
<tr>
<td></td>
<td>Tx-naive at entry – no</td>
<td>39.0 ± 7.4</td>
<td>24.5–53.5</td>
</tr>
<tr>
<td></td>
<td>Tx-naive at entry – N.S.</td>
<td>27.8 ± 3.3</td>
<td>21.3–34.3</td>
</tr>
<tr>
<td></td>
<td>GAF – yes</td>
<td>60.0 ± 0.0</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>GAF – no</td>
<td>38.8 ± 0.4</td>
<td>38.0–39.6</td>
</tr>
<tr>
<td></td>
<td>GAF – N.S.</td>
<td>56.4 ± 1.6</td>
<td>53.3–59.5</td>
</tr>
<tr>
<td></td>
<td>Employment/education – yes</td>
<td>49.3 ± 4.8</td>
<td>39.9–58.7</td>
</tr>
<tr>
<td></td>
<td>Combination therapy – N.S.</td>
<td>25.3 ± 4.8</td>
<td>15.9–34.7</td>
</tr>
<tr>
<td></td>
<td>EPI representativeness – yes</td>
<td>24.2 ± 4.0</td>
<td>16.4–32.0</td>
</tr>
<tr>
<td></td>
<td>EPI representativeness – no</td>
<td>49.9 ± 4.6</td>
<td>40.9–58.9</td>
</tr>
</tbody>
</table>

CI, Confidence interval; GAF, Global Assessment of Functioning Scale; EPI, epidemiological; Tx, treatment; N.S., not specified; N.A., not available.

**Good outcome**

Six predictors of good outcome were significant in univariate analysis and were then included in multivariable analysis (Table 3). Of these, having combination therapy, a non-representative sample and being from a developing country were associated with higher rates of good outcome (Table 4).

**Poor outcome**

Two predictors of poor outcome were significant in univariate analysis and were then included in multivariable analysis (Table 3). The use of typical antipsychotics and being treatment-naive at study entry were associated with higher rates of poor outcome (Table 4).
Readmission

The predictors selected for use in this model included follow-up duration, country of origin (developed versus developing) and patient status at start of study (in-patients versus combination in-/out-patients). When tested in a multivariable fashion, none of these effects reached statistical significance.

Employment/education

Three predictors were significant in univariate analysis and present in sufficient studies to be tested simultaneously in multivariable analysis (Table 3). Of these, combination therapy and a non-representative sample were associated with higher rates of employment/education (Table 4).

GAF

Six predictors of GAF were significant in univariate analysis and present in sufficient studies to be tested simultaneously in multivariable analysis (Table 3). Only being treatment-naive at study entry was associated with a higher GAF (better outcome) (Table 4).

Analyses by stratification

Stratification (by sample representativeness, diagnosis, study design and follow-up duration) was carried out for calculation of outcomes and for multivariable modeling of predictors. Outcome rates by strata (Table 2) revealed that some worst outcome domains were associated with: representative samples, longer study durations, study populations of primary diagnosis = schizophrenia, and prospective design. In multivariable analysis, stratification had little impact in examining predictors of outcome. This was most often because the strata were insufficiently powered for multivariable modeling, or for demonstrating any significant predictors or differences between strata.

DISCUSSION

Findings

Our analysis of FEP outcome studies revealed that the existing FEP literature is heterogeneous in its description of predictors and outcomes, making combination and comparison difficult. For studies that reported these outcomes, a good outcome was reported for 42% of the population, an intermediate outcome for 35%, and a poor outcome for 27%. There was no clear relationship between good/intermediate/poor outcome categories, as different studies reported each; the independence between studies and the reported outcomes may have contributed to predictors being both favorable and unfavorable contributors to outcome.

Few variables were found to be significant predictors of outcome. In summary, studies with non-epidemiologically representative samples or combination therapy were associated with better outcomes (good outcome, higher employment/education). In addition: (1) a developing country of origin was associated with good outcome; (2) use of typical (versus atypical or a combination of) neuroleptics and being treatment-naive at study entry were associated with poor outcome; and (3) being treatment-naive at entry was associated with higher GAF.

For the studies reporting rate of outcome, the rate of good outcome (42%) after a mean of 3 years’ follow-up was comparable to the 40% rate of improvement reported by Hegarty et al. (1994). Given the multiple definitions subsumed in ‘improvement’, it is likely that the ‘intermediate’ category reported in the current review overlaps to some degree with the ‘improvement’ of Hegarty et al. and, when considered with the good outcome group, would yield a higher rate than that of Hegarty et al. However, the absence of a uniform definition of outcome in the current analysis makes direct comparison between these results challenging. There are significant differences between Hegarty et al.’s meta-analysis and the current review, such as longer mean follow-up (6 years), greater percentage of chronic patients, and calculation of percentage improvement (which itself was broadly defined), that may have contributed to lower rates of improvement in their meta-analysis. The findings of our review are in keeping with the study hypothesis that analysis of similar outcome data for the FEP population would yield a higher rate of improvement.

Similar to the findings of Hegarty et al., follow-up duration was not a significant predictor of outcome. However, stratification by study duration does suggest some trends: with longer duration studies, there was a drop in rates of good outcome, an increase in rates of intermediate outcomes and a relatively minor
increase in poor outcome. Longer duration studies had increased rates of readmission and relapse, as could be expected, although this does not necessarily reflect their state of outcome or recovery. Finally, rates of functional recovery and mean GAF did not differ greatly according to study duration. Although our data demonstrate lower rates of good outcome in studies with longer follow-up, there is no proportionate increase in the rate of poor outcome, nor a decline in functional recovery and GAF. As such, our data do not suggest a clear pattern of progression in these studies.

There continues to be much ongoing debate about whether schizophrenia is a progressively deteriorating illness (Lieberman, 1999). If this was the case, a trend toward increasing percentages of patients with poor outcomes in studies with longer follow-ups would be expected. That this was not the case raises the possibility that those patients with poor outcomes may be apparent early in the course of illness and that the size of this group may be steady over time, reflecting an absence of progressive deterioration as suggested by other groups (Mason et al. 1996). It also brings to question what other factors may be contributing to course over time, for example housing, vocational and social issues that are not intrinsic components of the biology of psychosis. However, it must be acknowledged that these analyses do not address the possibility of patients moving into and out of the ‘poor outcome’ group, thus these issues require further study.

Samples classified as not being epidemiologically representative were associated with better outcomes. This may initially seem counter-intuitive; a representative sample would be expected to include all groups of prognoses and thus to have a better outcome than non-representative samples, which are usually biased toward sicker patients. However, our finding may reflect the possibility that outcomes can be inflated by a non-representative sample, for example in an academic setting where patients may come from a biased sample with greater family support, more education, and so on.

Medication variables (the use of typical medication and being treatment-naive at study entry) were primarily associated with poor outcome. The association with the use of typical neuroleptics may represent a time effect (i.e. typical neuroleptics more likely to be used in older studies associated with higher doses, less psychosocial rehabilitative programs, program differences such as lower threshold for hospitalization resulting in higher readmission rates, etc.). The association with being treatment-naive may reflect an issue of decreased accessibility to care and resources; however, being treatment-naive was associated with increased GAF (better). While it is difficult to explain these opposite findings, this may be further evidence of the impact of sampling effects and differing outcome definitions, possibly reflecting differences in setting or country of origin. The lack of association between pharmacotherapy and other outcomes may not contradict the existing literature associating medication with a lowered risk of relapse (Tauscher-Wisniewski & Zipursky, 2002). Instead, it may reflect the fact that most studies in the current analysis did have medication treatment, diminishing the detectable difference between groups. The significance of combination therapy as a positive predictor supports the importance of a psychosocial therapy component in addition to pharmacotherapy, and demonstrates the significance of biopsychosocial treatments in getting people back to work/school after an FEP. However, the current analysis cannot control for potential confounders associated with psychosocial therapies, such as type of intervention, impact on medication/follow-up, compliance, and phase of administration, which probably differed between studies and would have differential impact on outcome (Haddock & Lewis, 2005).

Given previous reports of variables such as age at onset, duration of untreated psychosis and diagnosis (e.g. schizophrenia versus schizophréniform or affective psychosis) being determinants of outcome, it was unexpected that none of these was identified as a significant predictor of outcome in this analysis. While this may be a function of differing definitions and decreased power (e.g. diagnosis criteria selected for schizophrenia-spectrum, diminishing the variability and differences, and subgroups of diagnoses may have been too small to be powered to show a difference), it may also suggest that sampling and types of intervention may be more significant determinants of outcome than demographic factors. However, the
process of stratifying by diagnosis did suggest that populations with a primary diagnosis of schizophrenia may have worse outcomes than those with broader diagnostic inclusion criteria. Results from stratification also suggest that study methodology issues (e.g. sampling, design, duration) may have an impact on outcome rates; for example, prospective studies would be more naturalistic, thus potentially associated with higher rates of non-compliance, drop-out and broader inclusion criteria admitting sicker patients, potentially increasing rates of worse outcomes.

Limitations

There are a number of significant limitations to this systematic review. Despite a literature search dating back to the 1960s, only studies published after the 1980s met criteria for analysis. In addition, despite the growing emphasis on the development of specialized FEP early intervention programs, there was a decrease in studies per annum after 2000, suggesting that the outcomes of these newer specialized programs may not be adequately reflected in this review.

A significant limitation of this review was the variability in definitions and study parameters. This divided studies into subgroups according to the definitions used, making comparison difficult and limiting the power to combine data and detect statistically significant differences. Few studies provided baseline measures from standardized clinical rating scales [e.g. the Brief Psychiatric Rating Scale (BPRS) and GAF], which would have permitted calculation of a rate of improvement for comparison. The variation in definitions was also reflected in the Hegarty meta-analysis, which broadly accepted that ‘patients considered as “improved” in follow-up had to have been described as recovered, in remission, well without residual symptoms, minimally or mildly symptomatic, improved without significant deficit, socially recovered, or working or living independently’ (Hegarty et al. 1994). However, given that all these definitions were translated into a rate of improvement, Hegarty et al. had a common definition for comparison of studies, whereas the current review had several of these definitions subsumed into ‘good’ and ‘intermediate’ categories by the original studies.

The absence of baseline and pre-morbid measures also precluded the analysis of an expected predictor of outcome, namely pre-morbid functioning. Given that baseline demographic and predictor sample characteristics (such as poor social relationships and level of unemployment) are frequently correlated with a significant proportion of outcome variance (Strauss & Carpenter, 1974b), and are essential for enabling the generalizability and replicability of the sample, it would be essential to have such measures (McGlashan et al. 1988).

This review eliminated retrospective follow-up (i.e. follow-back) studies even though this is a common design for follow-up studies, given its advantages (cost-effective permitting long follow-up) (Ram et al. 1992). However, follow-backs are vulnerable to missing and variable quality data, and cannot control for the quality and content of the past history data (McGlashan et al. 1988). The exclusion of follow-backs resulted in a reduction in the number of studies used, and loss of data from some interesting studies, often of long duration. These include the 15-year follow-up confirming the inverse correlation between duration of untreated psychosis (DUP) prior to first admission and long-term outcome (Bottlender et al. 2003), and some pivotal World Health Organization studies such as the International Study of Schizophrenia (ISOs) study that combined cohorts and demonstrated the initial 2-year course pattern as being the strongest predictor of 15-year outcome, with a rate of 50% favorable outcome for the group with schizophrenia (Harrison et al. 2001).

As in any review, an additional limitation includes those studies that may have been missed, particularly studies that may have been unpublished due to negative findings (mainly in drug trials). Given that the majority of included studies were observational studies (not trials) in which the reporting of all naturalistic outcomes would be expected, the issue of publication bias may not be as significant, although it is a potential limitation.

Finally, there may be an intrinsic limitation in using meta-analytic techniques for combining such heterogeneous literature to gain a larger sample and yield summary results. A meta-analytic approach permits a more objective appraisal of the evidence than traditional
narrative reviews (Egger et al. 1997). It can be a useful technique for providing a more precise and definitive answer when the results from individual studies disagree (Egger et al. 1998), and for exploring and sometimes explaining the heterogeneity between study results (Egger & Smith, 1997). However, it involves combination of data that must implicitly be somewhat comparable, and therein lie the main limitations of this review. An additional limitation is confounding, particularly intrinsic to observational designs, thus compounded in any attempt to combine observation studies. As the current review is based on an amalgamation of samples and is not individual-based, the restricted capacity to account for potential confounders (e.g. medication discontinuation and its impact on outcome) is an important limitation. Finally, the absence of a uniform definition of outcome limited the examination and reporting of an ‘effect size’ as is typically performed in a meta-analysis. As a result, this study has been designated a systematic review and not a meta-analysis, given that the heterogeneity of studies precluded the usual steps of a meta-analysis such as analytic techniques for estimates of effect, heterogeneity analysis and funnel plots.

Methodological issues in the existing literature and future research implications

The goal of this review, to combine results from existing studies, is novel for the FEP literature. It uses the power of a larger sample size, and thus adds some interesting findings to those previously reported. However, its limitations reflect those of the existing literature. This attempt at combining this literature emphasizes such limitations, drawing attention to areas for improvement.

As demonstrated in this review, defining the parts of the equation (i.e. the diagnosis and outcome) is central to clarifying the true outcome of schizophrenia. As discussed by Wing (1988), outcome itself remains to be defined in a universal and thus comparable way. In addition, how it is defined will determine rates of outcome (Warner, 1994), with higher rates of outcome when symptomatic remission is emphasized versus functional recovery (Robinson et al. 2005). Modern research has emphasized symptomatic outcome (Emsley, 1996; Bustillo et al. 1999). As discussed by Strauss & Carpenter (1974a) and demonstrated in this review, a diagnosis of schizophrenia based on symptom criteria alone is a weak predictor of outcome function; rather, outcome is not a single process but is composed of several semi-independent processes in different areas such as social relations, employment, symptoms and duration of hospitalization (Strauss & Carpenter, 1974b).

Liberman and colleagues (Liberman et al. 2002; Liberman & Kopelowicz, 2002a) have proposed that certain domains are central to defining recovery, including symptoms and functional status (vocations, independent living, and social relationships). A multi-dimensional definition of outcome is particularly important in the FEP population, where a greater percentage of patients would be expected to have at least some response to antipsychotic medication, resulting in inflated response rates, despite overall limited social functioning (Sheitman et al. 1997). Often the clinical and social/functional paths do not recover in a parallel fashion (Ciompi, 1980; Harding et al. 1987; Tohen et al. 2000; Liberman et al. 2002; Whitehorn et al. 2002) and should be evaluated separately in reporting outcome (Harrison & Mason, 1993).

To understand whether outcomes from schizophrenia will improve with early intervention and comprehensive approaches to early treatment, it will be important for future longitudinal outcome studies to incorporate standard design features that will enhance comparability across studies including: prospective follow-up of at least 2 years’ duration; inclusion of baseline measures; confirmation of diagnosis at least 1 year later; a large epidemiologically representative sample with in- and out-patients; a multi-dimensional model of outcome.
incorporating symptomatic/functional/personal elements measured at multiple time points; the use of standard and reliable scales for outcome measures; measures of potential determinants of outcome (e.g., treatment compliance, substance use, co-morbidity, pre-morbid functioning, cognitive status, etc.); and record of all interventions administered (pharmacologic, psychosocial), preferably following a standardized treatment algorithm that guides management and could be replicable and thus amenable to analysis in other centers.

CONCLUSIONS

In conclusion, this systematic review draws out the inconsistencies in the existing FEP outcome literature, while suggesting that incident samples of the schizophrenia-spectrum population may demonstrate higher levels of better outcome. It also suggests that while rates of good outcome may decrease in longer studies, rates of poor outcome do not necessarily increase with time as would be expected in a progressive deteriorating illness. Finally, the results emphasize sample representativeness and use of combination psychosocial/pharmacological therapy as influences of outcome. The heterogeneity in the outcome literature raises the question of whether schizophrenia truly is a group of diseases for which it is difficult to elucidate predictors of outcome (Riecher-Rossler & Rossler, 1998), or whether some significant portion of the variance is attributable to methodological variability and the use of multiple definitions of outcome. Research based on a multi-dimensional model of outcome can help to clarify this issue and lead us to understand the true outcome of schizophrenia given all our interventions. We need to be capable of discerning which patients do well so that we can understand why and thus how to optimize and ensure this through our detection and intervention programs.

ACKNOWLEDGMENTS

We thank Dr. T. Einarson who provided initial guidance in structuring the meta-analytic approach for this systematic review. Dr. Zipursky is supported by the Tapscott Chair in Schizophrenia Studies at the University of Toronto.

Dr. Menezes is supported by a training grant from the Canadian Institute for Health Research.

DECLARATION OF INTEREST

None.

REFERENCES


Systematic review of outcome of first-episode psychosis


