The prognosis of schizophrenia: A systematic review and meta-analysis with meta-regression of 20-year follow-up studies

Ida-Marie Molstrom, Julie Nordgaard, Annick Ufer-Parnas, Jonas Berge, Mads Gram Henriksen

Objective: The aim was to examine the general outcome of schizophrenia after 20 years or more.

Methods: Using the PRISMA guidelines, we conducted a systematic review and meta-analysis with meta-regression on long-term follow-up studies of schizophrenia up until April 21, 2021. We included prospective studies with at least 20 years of follow-up on patients with a diagnosis of schizophrenia, and the studies had to include face-to-face clinical evaluation. We examined outcome in three nested groups: ‘recovery’, ‘good or better’ (including also ‘recovery’), and ‘moderate or better’ (including also ‘recovery’ and ‘good or better’). We used random-effects meta-analysis and meta-regression to examine mean estimates and possible moderators.

Results: We identified 1089 records, which were screened by two independent researchers. 14 prospective studies (1991 patients) published between 1978 and 2020 were found eligible. The studies used a range of different scales and definitions for outcome, and some used the same definitions for different outcomes. To compare outcome across studies, we designed and applied a unified template for outcome definitions and cutoffs, based on earlier studies’ recommendations. Our meta-analysis found that 24.2 % had ‘recovered’ (n = 246, CI: 20.3–28.0 %), 35.5 % had a ‘good or better’ outcome (n = 766, CI: 26.0–45.0 %), and 59.7 % had ‘moderate or better’ outcome (n = 1139, CI: 49.3–70.1 %).

Conclusions: The results contribute to debunk the myth that schizophrenia inevitably has a deteriorating course. Recovery is certainly possible. Schizophrenia remains, however, a severe and complex mental disorder, exhibiting a limited change in prognosis despite >100 years of research and efforts to improve treatment.

1. Introduction

Schizophrenia has traditionally been associated with concepts of progression, relapse, and chronicity. The source of this view is usually ascribed to Kraepelin, who described dementia praecox as a disease with a chronic course and poor prognosis (Kraepelin, 1999). Yet, he also described that remission and recovery from schizophrenia were possible, and that heterogeneity was a marker of both the clinical presentation and the outcome of the disorder (Kendler, 2021). This view has been corroborated by longitudinal studies on schizophrenia, which have found a proportion of patients to have good outcome or recovery (McGlashan and Carpenter, 1988). However, substantial heterogeneity both between studies and within study groups exists, allowing noticeable variation in assessment of both course and outcome of schizophrenia. Methodological issues such as sample characteristics, diagnostic methods, follow-up assessments, and outcome measures have been highlighted as sources of this heterogeneity (Davidson and McGlashan, 1997; Jobe and Harrow, 2005).

In the last decades, there has been a growing interest in recovery and remission in schizophrenia (Andreasen et al., 2005; Huxley et al., 2021; Jääskeläinen et al., 2013; Lally et al., 2017; Warner, 2009). This has sparked efforts to clarify what recovery means. For example, Jääskeläinen et al. (2013) define recovery as both clinical and social remission, requiring that either clinical or social remission has lasted for at least two years. This definition is largely consistent with earlier proposed definitions (see e.g., Bleuler, 1978; Gross and Huber, 1986).
Remission in Schizophrenia Working Group (RSWG) has made a standardized definition for clinical remission, i.e., an improvement in core symptoms to a low-mild intensity level, where they no longer interfere significantly with behavior (Andreasen et al., 2005). Warner describes how social environment may have a profound effect on outcome of psychosis, and that social remission can be defined as economic and residential independence with low social disruption (Huxley et al., 2021).

Jääskeläinen et al. (2013) conducted a meta-analysis on follow-up studies of schizophrenia of varying durations with details on recovery and found a mean recovery rate of 16.4%, with no gender difference, and no change over time or between different follow-up periods, but also a higher recovery rate in low- and middle-income countries. Another meta-analysis by Hegarty et al. (1994) examined studies with a follow-up length of at least one year and found that 40.2% had good outcome after an average of 5.6 years. While reviews and meta-analyses on the very long-term outcome of schizophrenia have been published (Ajnakina et al., 2020, 2021; Angst, 1988; Carpenter and Kirkpatrick, 1988; Huxley et al., 2021; Joe and Harrow, 2005; Lang et al., 2013; McGlashan, 1988; Volavka and Vevera, 2018), they usually include studies of markedly different follow-up periods or, include schizophrenia in a broader group of psychotic disorders with better long-term outcome than schizophrenia (Periotogiannis et al., 2020). Based on the available literature, it is not possible to draw conclusions about the very long-term outcome of schizophrenia. So far, no systematic review or meta-analysis has focused exclusively on the very long-term outcome of schizophrenia.

The aim of our study was to shed light on this issue, i.e., to explore the very long-term outcome of schizophrenia. Consequently, we conducted the first systematic review and meta-analysis with meta-analysis of all prospective follow-up studies on schizophrenia, spanning at least 20 years.

2. Material and methods

2.1. Search strategy and selection criteria

Following the PRISMA guidelines (Page et al., 2021), on April 21, 2021, we conducted a systematic search of literature on long-term follow-up studies of schizophrenia of 20 years or more in PubMed, PsycInfo, and EMBASE, using the following search string:

((schizophrenia OR schizoaffective OR psychosis OR psychoses) OR (psychotic)) AND ((longitudinal) OR ((long-term) OR (follow-up)) OR (“follow-up”)) AND (((20 year*) OR (20-year*)) OR (twenty year*) OR (twenty-year*) OR ((21 year*) OR (21-year*)) OR (twenty-one year*) OR (twenty-one-year*) OR (twenty-two year*) ... continued until 40 year ... OR (very long term)) OR (very long-term))

To save space, we have inserted “… continued until 40 years … ” in the search string above. This signifies that we included all years from (“20 year*”) - in the various spellings - to (“forty-year*”) in the full search string that we used.

Finally, reference lists from systematic reviews, book chapters, and articles were screened for additional articles.

The protocol was registered on PROSPERO (CRD42021252124). The following inclusion criteria were used for study eligibility:

I. Peer reviewed prospective cohort studies, in English language. Studies may contain retrospective parts of the study, but they may not be solely based on retrospective reports.

II. Follow-up period of at least 20 years.

III. Baseline and follow-up assessment must include face-to-face clinical evaluation and not be restricted to questionnaires or register-based data.

IV. The study must provide information on the applied diagnostic method. Studies with diagnoses made retrospectively solely were excluded.

V. Diagnosis of schizophrenia, schizoaffective disorder, schizophrenomform disorder, or non-affective psychosis were included.

Studies without differentiation between types of psychosis, e.g., FEP cohorts, were excluded.

VI. Clearly defined outcome measures. Follow-up data on general, psychopathological, or social function outcome.

VII. Minimum follow-up sample size of 10 patients.

VIII. Sample is not restricted to children.

Two of the authors (I.M. and R.H.) independently screened the titles and abstracts, after settling on inclusion- and exclusion criteria with authors J.N., A.U.P., and M.G.H. Studies were included for full-text assessment in cases of doubt. The retrieved articles’ reference lists were also reviewed to further identify potentially relevant studies. Disagreement regarding inclusion was resolved through a meeting between all authors.

2.2. Data extraction

The following data was extracted: First-author, year of publication, study location, study design, setting (inpatient, outpatient, mixed), number of follow-up assessments, duration of follow-up, first-admission sample or mixed, sample size at baseline and last follow-up, gender, age at follow-up, diagnosis and diagnostic method, outcome measurements, mortality and lost to follow-up. If more than one diagnostic evaluation was made, this was reported in the table, and the earliest diagnosis was used in the analysis. We extracted clinical, social, and overall outcomes at the latest follow-up point (20+ years) as outcome data. The data were either extracted from articles or converted from tables and figures provided. Ambiguities were resolved by consensus among the authors.

2.3. Definition of outcomes

To compare outcomes across studies using heterogeneous and sometimes even inconsistent definitions or cutoffs, we constructed a unified template with standardized definitions for different outcomes, i.e., recovery, good, moderate, and poor outcome. Our definitions were as far as possible based both on definitions in the original studies and on recommendations from other studies (Andreasen et al., 2005; Bleuler, 1978; Gross and Huber, 1986; Harrison et al., 2001; Helgason, 1990; Huxley et al., 2021; Jääskeläinen et al., 2013; Kua et al., 2003; Newman et al., 2012; Ogawa et al., 1987; Sartorius et al., 1996; Warner, 2009) (see Table 1). We then superimposed these outcome definitions onto the reviewed studies. A similar approach was applied in the meta-analysis by Hegarty et al. (1994) and Jääskeläinen et al. (2013).

We defined recovery as both clinical and social remission, and one of these must have lasted for at least two years (Jääskeläinen et al., 2013). Good outcome is defined as the presence of clinical and/or social remission, which is in line with prior definitions by Bleuler (1978), Gross and Huber (1986) and Hegarty et al. (1994). We used the definition for clinical remission constructed by the RSWG, as “no or mild symptoms”, or as a score of mild or less on the Positive and Negative syndrome scale (PANSS), the Scale for Assessment of Positive Symptoms (SAPS), the Scale for Assessment of Negative Symptoms (SANS) or the Brief Psychiatric Rating Scale (BPRS) (Andreasen et al., 2005). Social remission is defined as good or very good social function, which can be evaluated from a range of different outcome measurements; full employment, normal social life, self-supportive, scores of good or very good social function on WHO Disability Assessment Schedule (WHODAS), or the Social and Occupational Functioning Assessment Scale (SOFAS) (Helgason, 1990; Kua et al., 2003; Ogawa et al., 1987; Warner, 2009). The Global Assessment of Functioning Scales (GAF/GAF) were also included with a cut-off of >60 (Endicott et al., 1976; Harrison et al., 2001; Sartorius et al., 1996). Moderate outcome is defined as intermediate or moderate symptoms and/or moderately impaired social function, e.g., part time employment. Poor outcome is defined as severe or unstable symptoms and/or poor social function, e.g., living isolated or hospitalized (Bleuler, 1978; Gross and Huber, 1986; Helgason, 1990; Kua et al.,
| Recovery defined as: | Good outcome: Good general, clinical or social outcome defined as one or more of the following: | Moderate outcome: Moderate social, clinical or general outcome defined as one or more of the following: | Poor outcome: Poor social, clinical or general outcome defined as one or more of the following: |
|----------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| General outcome: Both clinical and social remission. One of these for at least 2 years. (Jaaskelainen et al., 2013). E.g.: “Complete remission”, monophasic or polyphasic illness course with ≥3 years complete recovery after psychotic phase, and full employment. (Gross and Huber, 1986). Recovered “end state” (≥5 years): Full employment, reassumed social roles. Currently no psychotic symptoms except for some eccentricity or symptom residues. (Bleuler, 1978). Returned to previous social functioning, independent social life, maintains a normal family life and no positive symptoms for the last 5 years. (Ogawa et al., 1987). | Mild “end state”: Maintains a sensible conversation, normal general behavior, illness not immediately obvious. Schizophrenic symptoms do exist. (Bleuler, 1978). Complete remission (Bonn Criteria for Psychopathological outcome. Gros and Huber, 1986; Marneros et al., 1989). GAS > 60: No or mild symptoms; superior functioning to some difficulty functioning. (GAF/GAS Scales. Endicott et al., 1979; Harrison et al., 2001; Sartorius et al., 1996). “Improved”: To remission, well without residual symptoms, minimally or mildly symtomatic, improved without significant deficit, socially recovered, or working or living independently. (Hegarty et al., 1994). | Moderately severe “end state”: As severe “end state”, but in one aspect or another better preserved, e.g. could communicate, could work in some form, could create fluctuating contact to others. (Bleuler, 1978). “Non-characteristic residua” (Bonn Criteria for Psychopathological outcome. Gross and Huber, 1986; Marneros et al., 1989). GAS 51-60: Moderate symptoms; difficulty functioning. (GAF/GAS Scales. Endicott et al., 1976; Harrison et al., 2001; Sartorius et al., 1996). | Severe “end state”: Unable of normal conversation. Not working or very limited work. Indifferent to their surroundings. Requires constant care. (Bleuler, 1978). “Characteristic residua” (Bonn Criteria for Psychopathological outcome. Gross and Huber, 1986; Marneros et al., 1989). GAS 50: Serious symptoms; impaired function or in need constant supervision. (GAF/GAS Scales. Endicott et al., 1976; Harrison et al., 2001; Sartorius et al., 1996). |
| Clinical outcome: Clinical remission: Low-mild symptom intensity level, which (absent/mild/ borderline symptoms) do not influence an individual’s behavior. (RSWG (Andreasen et al., 2005)). PANSS ≤ 3 (positive, negative and disorganized symptoms). SAPS and SANS ≤ 2. BPRS ≤ 3. (RSWG (Andreasen et al., 2005)). GAF-S > 60 | Intermediate or moderate symptoms (Helgason, 1990). GAF-S 51-60 | Severe or unstable symptoms (Ogawa et al., 1987). GAF-S ≤ 50 |
| Social outcome: Social remission: Good or very good social function. E.g. self-supportive, fully employed, normal relationships, has friends and is socially active. (Helgason, 1990; Kua et al., 2003; Ogawa et al., 1987; Warner, 2009). GAF-D > 60 WHODAS 0-1: Excellent or good function, with no or minor dysfunction. (World Health Organisation (WHO), 1988). SOFAS = 60: Superior function to functioning well. (Goldman et al., 1992). Loss of productive time < 10 %. (Newman et al., 2012). | Moderately impaired social function. E.g. semi-self-supportive, part time employment, or few social interactions, takes part irregularly. (Helgason, 1990; Kua et al., 2003; Ogawa et al., 1987) GAF-D 51-60 WHODAS 2: “Fair” function. (World Health Organisation (WHO), 1988). SOFAS 51-60: Moderate dysfunction. (Goldman et al., 1992). Loss of productive time 11–50 %. (Newman et al., 2012). | Poor very poor social function. E.g. Maladjusted, living isolated, hospitalized, no friends, not working. (Helgason, 1990; Kua et al., 2003; Ogawa et al., 1987). GAF-D ≤ 50 WHODAS 3-5: Serious or very serious or maximum dysfunction. (World Health Organisation (WHO), 1988). SOFAS ≤ 50: Severe dysfunction or inability to function. (Goldman et al., 1992). Loss of productive time ≥ 50 %. (Newman et al., 2012). |

GAF: Global Assessment of Functioning Symptoms/Disability (American Psychiatric Association, 1994).
GAS: Global Assessment Scale Evaluation over 1 month (Endicott et al., 1976).
WHODAS: WHO Disability Assessment Schedule (World Health Organisation (WHO), 1988).
SOFAS: The Social and Occupational Functioning Assessment Scale, subscale of the Global Assessment of Functioning (Goldman et al., 1992).
RSWG: Remission in Schizophrenia Working Group (Andreasen, 2005).
PANSS: Scale for Assessment of Positive Symptoms (Andreasen and Olsen, 1982).
SAPS: Scale for Assessment of Negative Symptoms (Andreasen, 1982).
BPRS: The Brief Psychiatric Rating Scale (Andreasen, 2005; Overall and Gorham, 1962).
* Criteria for clinical and social remission are the same as for good outcome, but recovery requires both clinical and social remission.
Cut-offs on the scales GAF/GAS, WHODAS and SOFAS are visible in Table 1. Newman et al. (2012) evaluated “loss of productive time”, which is also noted with cut-offs in Table 1. We have not included treatment requirements in the definitions, nor have we included a time criterion for good, moderate or poor outcome, because this is rarely stated. For more details see Table 1.

All general, clinical, and social outcomes from the included studies were extracted as well as their use of definitions and cut offs. Outcomes were reported as means (%). Each outcome was evaluated in accordance to Table 1, and a decision by the authors in consensus was made whether to include, exclude or moderate the outcome as to meet our general standardization. The decisions are reported in Table 3. If a study reported more than one result for an outcome category, a mean was calculated for the meta-analysis.

In the meta-analysis, we examined the long-term outcome using three variables: ‘recovery’, ‘good or better’ (including also ‘recovery’), and ‘moderate or better’ (including also ‘good or better’ and ‘recovery’). Poor outcome is the inverse of ‘Moderate of better’, i.e., poor outcome refers to study participants, who did not meet the criteria for ‘Moderate or better’ outcome.

Fig. 1. Nested groups for meta-analysis: ‘Recovery’, ‘Good or better’ (including also ‘Recovery’), and ‘Moderate or better’ (including also ‘Good or better’ and ‘Recovery’). Poor outcome is the inverse of ‘Moderate of better’, i.e., poor outcome refers to study participants, who did not meet the criteria for ‘Moderate or better’ outcome.
Table 2
Overview of the 14 included studies.

| Author(s), publication year, location | Study type | N baseline/follow-up (% women) | Diagnosis | Diagnostic method | Year/type of enrollment | Year of follow-up/(study duration) | Age at follow-up (years) | Mortality (%) | Attrition (%) |
|--------------------------------------|------------|--------------------------------|------------|-------------------|------------------------|-----------------------------------|--------------------------|--------------|--------------|
| Bleuler, 1972/Burgholzli, Zurich     | ‘Real-time’ prospective study. ‘Catch-up’ prospective study. | 208/138 (51 %) | SZ          | Criteria based on E. Bleuler but requires psychosis. Criteria according to E. Bleuler | 1942-43. First-and readmission | 1965 (22.5 years)             | 42.5          | 34 %         | 33.7 %       |
| Ciompi, 1980/The Lausanne Investigations | ‘Catch-up’ prospective study. Single follow-up. | 1642/289 (68 %) | SZ          | SZ                  | 1900–62. First- and readmission | 1964–69 (36.9 years)              | 74           | 75 %         | 82.4 %       |
| Huber et al., 1986; Gross and Huber, 1986/The Bonn Study | ‘Real-time’ prospective study. 2 follow-up assessments. | 758/502 (58 %) | SZ | Criteria according to K. Schneider and M. Bleuler. DSM-I, DSM-II and DSM-III (retrospectively re-diagnosed) | 1945-59. 67 % first admission. | 1955-1959. Long term hospitalized, chronic cases. | 1980–1982 (32 years) | 61 | 23.7 % 30.5 % |
| Harding et al., 1987/The Vermont Longitudinal Study | ‘Real-time’ prospective study. 2 follow-up assessments. | 118/82 (50 %) | SZ (selected chronic group). | SZ                  | 1955–1959. Long term hospitalized, chronic cases. | 1980–1982 (32 years) | 61 | 23.7 % 30.5 % |
| Ogawa et al., 1987/Gunma University Hospital Japan | ‘Real-time’ prospective study. | 140/98 (54 %) | SZ          | ICD-9               | 1958–62. 79 % first admission. | (24 years)              | 48           | 17.9 % 25.0 % |
| Marneros et al., 1989/Cologne Study | ‘Catch-up’ prospective study. Single follow-up. | 97/97 (43 %) | SZ (subgroup, compares to SA) | DSM-III/IIR       | 1950–79. First and readmission. | (23 years) | 49 | 0 % |
| Opjordsmoen, 1986, 1991/Oslo | ‘Real-time’ prospective study. 2 follow-up assessments. | 94/70 (54 %) | SZ (subgroup, compares to other psychoses) | ICD-9 and DSM-III | 1946–48. First admission | Around 1986 (30 years) | 61 | 33.7 % 25.5 % |
| Helgason, 1990/Reykjavik Iceland | ‘Real-time’ prospective study. 2 follow-up assessments. | 107/82 (46 %) | SZ          | ICD-8               | 1966–67. First admission/First contact | 1987 (21 years) | 55 | 21.5 % 23.4 % |
| Kua et al., 2003/Singapore | ‘Real-time’ prospective study. | 402/216 (30 %) | SZ          | ICD-9               | 1975. First admission. | 1995 (20 years) | 43 | 14.7 % 46.3 % |
| Thara, 2004/Madras India | ‘Real-time’ prospective study. | 90/61 (46 %) | SZ          | ICD-9               | 1981-82. First contact. | 2001–20 (20 years) | 44.5 | 17.8 % 32.2 % |
| Grossman et al., 2008; Harrow and Jobe, 2010/Chicago Follow-Up Study | ‘Real-time’ prospective follow-up study. 6 follow-up assessments. | 51/43 (33 %) | SZ (subgroup, compares to other psychoses and non-psychotic group.) | DSM-III | 1978. 41 % first admission. (Private hospital) | 1998 (20 years) | 43 | 15.7 % |
| Newman et al., 2012/Alberta Canada | Retro- and prospective study | 128/81 (57 %) | SZ          | DSM-II              | 1963. First admission. | 1997 (29 years) | 62 | 36.7 % 36.7 % |
| Kotov et al., 2017; Velikhov et al., 2017/Suffolk County Mental Health Project | ‘Real-time’ prospective study. 6 follow-up assessments. | 248/175 (40 %) | SZ (subgroup, compares to other psychoses and non-psychotic group.) | DSM-IV | 1990–1995. First admission. | 2010–2015 (20 years) | 49 | 12.6 % 29.4 % |
| Cechnicki et al., 2020; Cechnicki et al., 2018; Cichocki et al., 2015/Cracow | ‘Real-time’ prospective study. 3 follow-up assessments. | 80/57 (40 %) | SZ          | DSM-III, DSM-IV | 1985–88. First admission. | 2005–08 (20 years) | 48 | 6.3 % 28.8 % |

SZ: Schizophrenia. SA: Schizoaffective disorder. 
ICD-8/9/10: International Classification of Diseases, 8th edition/9th edition/10th edition. 
DSM-I/II/III/IIR/IV: Diagnostic and statistical manual of mental disorders 1st edition/2nd edition/3rd edition/3rd edition revised/4th edition.
| Author(s) | Outcomes of the 14 included studies. |
|-----------|-------------------------------------|
| Bleuler, 1972 | Recovery: Recovery 'end state' (five years). Good: Recovery and mild 'end stage'. Moderate: Moderately severe 'end stage'. Poor: Severe, uncertain or unstable 'end stage'. |
| Glompi, 1980 | Recovery: Bleuler's recovery. Good: Recovery and mild 'end stage'. Moderate: Intermediate 'end stage'. Poor: Severe, uncertain or unstable 'end stage'. |
| Huber et al., 1989 | Recovery: 'Full remission', monophasic or polyphasic course of illness with complete recovery ≥3 years, and fully employed. Good: a) 'Full remission'. b) Fully employed at previous/lowest level. Moderate: a) 'Noncharacteristic residuals', b) Limited ability to work. Poor: a) 'Characteristic residuals', b) Incapable of working. |
| Gross and Huber, 1986 | Good: a) Global Assessment Scale (GAS) 61–100. b) Employed in past year. c) Displayed slight or no symptoms. d) Met with friends every week or 2. Moderate: GAS 31–60. Poor: GAS 0–30. |
| Marneros et al., 1989 | Good: a) Bonn Criteria for Psychopathological outcome - “full remission” for >3 years. b) GAS > 90 (GAS > 60 added from article by the authors). Moderate: a) Bonn Criteria for Psychopathological outcome - “non-characteristic residuals”, b) GAS 51–90 (GAS 51-60 added by the authors). c) WHODAS score 0.51–2. Poor: a) Bonn Criteria for Psychopathological outcome - “characteristic residuals”, b) GAS < 50, c) WHODAS score >2.0. |
| Ojjojordemoen, 1986, 1991 | Recovery: GAS > 90. “Full recovery”. Good: GAS > 60. Moderate: GAS 51–60. Poor: GAS < 50. |
| Helgason, 1990 | Good: a) No/minor symptoms + no treatment/treatment outside hospital. b) Normal relationships, has friends and is socially active. c) Full time employment or >6 mo/yr. Moderate: a) Obvious symptoms of psychopathology. Admission <2 months/year. b) Few close friends, takes part irregularly. c) Works sometimes. Poor: a) Severe symptoms, hospitalized or sheltered living >2 months/yr. b) No social interactions, tendency to isolation/isolated. c) No employment. |
| Kua et al., 2003 | Good: Patient not receiving treatment, well and working. Moderate: Patient not receiving treatment and not working or receiving out-patient treatment and working. Poor: Patient receiving treatment and not working or receiving in-patient treatment. |
| Thara, 2004 | Good: a) GAF-S > 60, b) GAF-F > 60. |
| Gousman et al., 2008; Harrow and Jobe, 2010 | Recovery: LKP 1–2 and SCS >2 - absence of major symptoms, adequate social function, no rehabilitation for 1 year. |
| Newman et al., 2012 | Good: a) WHODAS 0–1. b) SOFAS >70. c) Loss of productive time <10 %. Moderate: a) WHODAS 2. b) SOFAS 51–70. c) Loss of productive time 11–50 %. Poor: a) WHODAS 3–5. b) SOFAS 1–50. c) Loss of productive time >50 %. |
| Kottov et al., 2017; Velthorst et al., 2017 | Good: a) Preserved social function, b) GAF ≥ 60. Moderate: Moderately impaired social function. Poor: Severely/profoundly impaired social function. |

### Table 3

| Author(s) | Outcome criteria used in studies | Outcome | Recovery | Good | Moderate | Poor | Our evaluation for inclusion in meta-analysis (MA) |
|-----------|---------------------------------|---------|---------|------|---------|------|---------------------------------|
| Bleuler, 1972 | Recovery: Recovery 'end state' (five years). Good: Recovery and mild 'end state'. Moderate: Moderately severe 'end state'. Poor: Severe, uncertain or unstable 'end state'. | 20 % | 53 % | 24 % | 24 % | All outcomes included. |
| Glompi, 1980 | Recovery: Bleuler's recovery. Good: Recovery and mild 'end stage'. Moderate: Intermediate 'end stage'. Poor: Severe, uncertain or unstable 'end stage'. | 26.6 % | 48.7 % | 23.9 % | 27.4 % | All outcomes included. |
| Huber et al., 1989 | Recovery: 'Full remission', monophasic or polyphasic course of illness with complete recovery ≥3 years, and fully employed. Good: a) 'Full remission'. b) Fully employed at previous/lowest level. Moderate: a) 'Noncharacteristic residuals', b) Limited ability to work. Poor: a) 'Characteristic residuals', b) Incapable of working. | 21.0 % | a) | 43.2 % | a) | 34.7 % | All outcomes included. Moderate and poor outcome included due to unacceptable cut-offs. |
| Gross and Huber, 1986 | Good: a) Global Assessment Scale (GAS) 61–100. b) Employed in past year. c) Displayed slight or no symptoms. d) Met with friends every week or 2. Moderate: GAS 31–60. Poor: GAS 0–30. | – | a) | 60 % | 40 % | 0 % | Good outcome included as mean score of a-d. Moderate and poor outcome included due to unacceptable cut-offs. |
| Marneros et al., 1989 | Good: a) Bonn Criteria for Psychopathological outcome - “full remission” for >3 years. b) GAS > 90 (GAS > 60 added from article by the authors). Moderate: a) Bonn Criteria for Psychopathological outcome - “non-characteristic residuals”, b) GAS 51–90 (GAS 51-60 added by the authors). c) WHODAS score 0.51–2. Poor: a) Bonn Criteria for Psychopathological outcome - “characteristic residuals”, b) GAS < 50, c) WHODAS score >2.0. | 32 % | a) | 57 % | a) | 25 % | Good and moderate outcome excluded due to unacceptable cut-offs. Comparable outcomes added in ( ), which are used in MA. Moderate and poor outcome excluded due to unacceptable cut-offs. |
| Ojjojordemoen, 1986, 1991 | Recovery: GAS > 90. “Full recovery”. Good: GAS > 60. Moderate: GAS 51–60. Poor: GAS < 50. | 6 % | 17 % | 23 % | 60 % | All outcomes except recovery outcome included due to lack of time criterion. All outcomes included. |
| Helgason, 1990 | Good: a) No/minor symptoms + no treatment/treatment outside hospital. b) Normal relationships, has friends and is socially active. c) Full time employment or >6 mo/yr. Moderate: a) Obvious symptoms of psychopathology. Admission <2 months/year. b) Few close friends, takes part irregularly. c) Works sometimes. Poor: a) Severe symptoms, hospitalized or sheltered living >2 months/yr. b) No social interactions, tendency to isolation/isolated. c) No employment. | – | a) | 29 % | a) | 50 % | 21 % | Moderate and poor outcome excluded due to unacceptable cut-offs. |
| Kua et al., 2003 | Good: Patient not receiving treatment, well and working. Moderate: Patient not receiving treatment and not working or receiving out-patient treatment and working. Poor: Patient receiving treatment and not working or receiving in-patient treatment. | – | 28.3 % | 37.0 % | 34.7 % | All outcomes included. |
| Thara, 2004 | Good: a) GAF-S > 60, b) GAF-F > 60. | – | a) | 77.1 % | b) | 73.8 % | All outcomes included. |
| Gousman et al., 2008; Harrow and Jobe, 2010 | Recovery: LKP 1–2 and SCS >2 - absence of major symptoms, adequate social function, no rehabilitation for 1 year. | 28 % | – | – | – | Recovery moved to good outcome in MA, since the time criterion is not 2 years. |
| Newman et al., 2012 | Good: a) WHODAS 0–1. b) SOFAS >70. c) Loss of productive time <10 %. Moderate: a) WHODAS 2. b) SOFAS 51–70. c) Loss of productive time 11–50 %. Poor: a) WHODAS 3–5. b) SOFAS 1–50. c) Loss of productive time >50 %. | – | a) | 25.0 % | a) | 57.1 % | All outcomes included except good b) and moderate b) due to unacceptable cut-offs. |
| Kottov et al., 2017; Velthorst et al., 2017 | Good: a) Preserved social function, b) GAF ≥ 60. Moderate: Moderately impaired social function. Poor: Severely/profoundly impaired social function. | – | a) | 3 % | 24 % | 73 % | All outcomes included. |
2.5. Meta-analysis

We performed a series of meta-analyses using proportional random-effects models with inverse-variance weighting. Random-effects models were used because of the differences in diagnostic procedures, cohort recruitment, demographies, treatment availability and other aspects, which makes it reasonable to assume that there will be a high degree of heterogeneity between studies. Previous systematic reviews and meta-analyses of similar research questions have indeed shown a high degree of heterogeneity (Boonstra et al., 2012; Hegarty et al., 1994; Jääskeläinen et al., 2013). Heterogeneity was described as I², which is a recommended transformation of the calculated Q. Values of I² range from 0 % to 100 %, reflecting the proportion of the total variation across studies beyond chance. The value of 25 % describes low, 50 % moderate, and 75 % high heterogeneity. As a sensitivity analysis, we performed an influence analysis for each of the three main analyses by running each analysis multiple times, each time excluding one of the studies. This allowed us to assess whether the results were robust against excluding particular studies (Borenstein, 2009).

2.6. Meta-regression-analysis

We also explored possible sources for heterogeneity both in methodology and by performing meta-regression analyses on possible moderators, which have been suggested or examined in earlier studies as contributors to heterogeneity in outcomes between studies, e.g., changes in outcome over time, or changes in outcome per diagnostic methodology (Harrison et al., 2001; Hegarty et al., 1994; Jääskeläinen et al., 2013; Menezes et al., 2006). We used the same analysis design as in the main analyses, adding each of the variables as a moderator in the model. We thus performed 15 meta-regression models (five moderator variables multiplied with three outcome measures). In cases in which the meta-regression models showed statistically significant effects of the moderator variables at the p < 0.05 level, subgroup analyses were performed to obtain proportion estimates for presentation. Influence analyses were also performed to ascertain that the results were not contingent on a particular study. In cases with more than one moderator variable showing a statistically significant association with a particular outcome measure, all the statistically significant moderator variables were entered in another meta-regression model to assess what associations remained substantial after this adjustment.

The analysis related to change over time was conducted by examining the studies according to the enrollment year, which was entered in the analyses as a continuous variable. The mean age at follow-up for the study sample was compared by creating two groups (age ≤ 50, age > 50), based on the pooled mean age being 52.17. The effect of the geographical placement of the study was examined by dividing the studies into the geographical groups of Europe, North America, or Asia. We also analyzed studies using preoperative diagnostic criteria (i.e., E. Bleuler’s criteria, M. Bleuler’s criteria, ICD-8, DSM-III-R, and ICD-10, DSM-IV). All statistical analysis was performed using R 4.0.2 (R Core Team, 2013) and the metafor package (Viechtbauer, 2010).

2.7. Attrition

In order to address the possible effects of the high attrition rates in the included studies, we performed a series of meta-regressions for each of the outcomes using attrition rate as an independent variable in the models.

3. Results

3.1. Study characteristics

14 studies published between 1978 and 2020 were eligible for inclusion (see Fig. 2 for flowchart; studies excluded after full-text assessment are presented in the Supplementary Material). Study characteristics of the included studies are presented in Table 2.

| Author(s) | Outcome criteria used in studies | Outcome measures | Our evaluation for inclusion in meta-analysis (MA) |
|-----------|----------------------------------|------------------|--------------------------------------------------|
| Gecelnick et al., 2018; 2020; Cichocki et al., 2015 | Recovery: a) BPRS ≤ 2. b) GAS ≥ 60. Good: Very good/good social function. Moderate: Satisfactory social function. Poor: Poor/very poor social function. | a) 21 % b) 56 % | Recovery a) and b) moved to good outcome in MA due to lack of time criterion for recovery. |

MA: Meta-Analysis.
GAS: Global Assessment Scale. Evaluation over 1 month (Endicott et al., 1976).
ESAS: Eguma’s Social Adjustment Scale (Eguma, 1962).
WHODAS: WHO Disability Assessment Schedule (World Health Organisation (WHO), 1988).
GAF: Global Assessment of Functioning Symptoms/Disability (American Psychiatric Association, 1994).
BPRS: The Brief Psychiatric Rating Scale (Overall and Gorham, 1962).
LKP: Levenstein, Klein, and Pollack scale (Levenstein et al., 1966).
SCS: Strauss-Carpenter Scale (Straus and Carpenter, 1972).
SOFAS: The Social and Occupational Functioning Assessment Scale is a subscale of the Global Assessment of Functioning (GAF).

"Moderate or better". Poor outcome is defined as the inverse of the variable 'moderate or better' (see Fig. 1 for details).
**Recovery**

| Author(s) and Year | N   | Recovery | Proportion [95% CI] |
|--------------------|-----|----------|---------------------|
| M. Bleuler 1972   | 138 | 28       | 0.20 [0.14, 0.27]   |
| Ciompi 1980       | 289 | 78       | 0.27 [0.22, 0.32]   |
| Huber et al. 1980 | 502 | 110      | 0.22 [0.18, 0.26]   |
| Ogawa et al. 1987 | 98  | 30       | 0.31 [0.21, 0.40]   |

RE Model (Q = 5.7, df = 5, p < .0001, I² = 12.9%)

**Good or better outcome**

| Author(s) and Year | N   | Good | Proportion [95% CI] |
|--------------------|-----|------|---------------------|
| M. Bleuler 1972   | 138 | 73   | 0.53 [0.45, 0.61]   |
| Ciompi 1980       | 289 | 141  | 0.49 [0.43, 0.55]   |
| Huber et al. 1980 | 502 | 196  | 0.39 [0.35, 0.43]   |
| Harding et al. 1987 | 82  | 47   | 0.57 [0.47, 0.68]   |
| Ogawa et al. 1987 | 98  | 46   | 0.47 [0.37, 0.57]   |
| Marreros et al. 1989 | 97  | 14   | 0.14 [0.07, 0.21]   |
| Opjordsmoen 1991 | 70  | 12   | 0.17 [0.08, 0.26]   |
| Helgason et al. 1990 | 82  | 18   | 0.22 [0.13, 0.31]   |
| Kua et al. 2003   | 216 | 132  | 0.61 [0.55, 0.68]   |
| Thara et al. 2004 | 61  | 28   | 0.46 [0.33, 0.58]   |
| Grossman et al. 2008 | 43  | 12   | 0.28 [0.15, 0.41]   |
| Newman et al. 2012 | 81  | 19   | 0.23 [0.14, 0.33]   |
| Velthorst et al. 2017 | 175 | 6    | 0.03 [0.01, 0.06]   |
| Cechnicki et al. 2020 | 57  | 22   | 0.39 [0.26, 0.51]   |

RE Model (Q = 596.1, df = 12, p < .0001, I² = 98%)

**Moderate or better outcome**

| Author(s) and Year | N   | Moderate | Proportion [95% CI] |
|--------------------|-----|----------|---------------------|
| M. Bleuler 1972   | 138 | 105      | 0.76 [0.69, 0.83]   |
| Ciompi 1980       | 289 | 211      | 0.73 [0.68, 0.78]   |
| Huber et al. 1980 | 502 | 354      | 0.71 [0.67, 0.75]   |
| Ogawa et al. 1987 | 98  | 70       | 0.71 [0.62, 0.80]   |
| Marreros et al. 1989 | 97  | 39       | 0.40 [0.30, 0.50]   |
| Opjordsmoen 1991 | 70  | 28       | 0.40 [0.29, 0.51]   |
| Helgason et al. 1990 | 82  | 61       | 0.74 [0.65, 0.84]   |
| Kua et al. 2003   | 216 | 148      | 0.69 [0.62, 0.75]   |
| Newman et al. 2012 | 81  | 38       | 0.47 [0.36, 0.58]   |
| Velthorst et al. 2017 | 175 | 47      | 0.27 [0.20, 0.33]   |
| Cechnicki et al. 2020 | 57  | 38       | 0.67 [0.54, 0.79]   |

RE Model (Q = 214.6, df = 10, p < .0001, I² = 95.3%)

---

Fig. 3. Forest plots for the nested outcome groups of ‘recovery’, ‘good or better’, and ‘moderate or better’.
et al., 1989; Oppjordsmoen, 1986). Eight studies used pre-operational diagnostic criteria, and six studies used operational diagnostic criteria. The study by Harding et al. (1987) enrolled from a selected, chronic group of patients with schizophrenia receiving specialized rehabilitation. The Chicago follow-up study recruited participants from a private hospital (Grossman et al., 2008). Six studies included only first admission or first contact patients. Mortality rates ranged from 6 % (Cechnicki et al., 2020) to 75 % (Ciompi, 1980) (the latter focused on late phase schizophrenia and had an average age at follow-up of 75 years).

3.2. Outcome

Our assessment of the outcomes for each study is presented in Table 3, where the studies' own definitions and scales used to measure and define outcomes also can be found. Several studies used different outcome definitions and cutoffs, and sometimes the same definition was used for different outcomes, e.g., Cechnicki et al. (2016) used the same definition for recovery as Kua et al. (2003), Oppjordsmoen (1991), and Harding et al. (1987) used for good outcome (i.e., GAS > 60).

3.3. Quality assessment

Risk of bias for each study and quality assessment are presented in Table 1 of the Supplementary material. Two studies were considered of good quality, receiving a rating of 7 (Grossman et al., 2008; Kotov et al., 2017), whereas the 12 remaining studies were considered of fair quality, receiving ratings of 5 or 6.

3.4. Meta-analysis

Four studies, comprising a total of 1027 study participants, included a comparable ‘recovery’ outcome measure, which was reported in 24.2 % (95 % confidence interval: 20.3–28.0 %) of the participants. ‘Good or better’ outcome was assessed in all 14 studies, including a total of 1919 study participants, and was reported in 35.5 % of the participants (95 % CI: 26.0–45.0 %). ‘Moderate or better’ outcome was assessed in 11 studies, comprising 1805 study participants, and was reported in 59.7 % of the participants (95 % CI: 49.3–70.1 %). Inversely, this also means that 40.3 % of the participants had poor outcome. The results of the three main analyses are summarized in forest plots (Fig. 3). As expected, we found the estimates highly heterogeneous, except for the recovery group which may be due the fact that only four studies were included (‘recovery’: $I^2 = 12.6 \%$, ‘good or better’: $I^2 = 98.0 \%$, and ‘moderate or better’: $I^2 = 95.3 \%$). Data included in the meta-analysis is provided in Supplementary Table 4.

3.5. Meta-regression analysis

The moderator variables ‘enrollment year’ and ‘age group’ did not show any statistically significant associations with any of the outcome variables. The moderator variable ‘geographical location’ did not show statistically significant associations with the outcome variables ‘recovery’ or ‘good or better’ (Table 4). The moderator variable ‘diagnostic method’ did not show a significant association with ‘recovery’. Yet, patients in studies using operational diagnostic criteria were significantly less likely to report ‘good or better’ outcome compared to studies using pre-operative diagnostic criteria (23.7 % vs. 44.0 %, mean difference 20.3 percentage points, 95 % CI: 4.2–36.5, $p = 0.014$). This was also the case for ‘moderate or better’ outcome (43.0 % (operational diagnostic criteria) vs. 69.3 % (pre-operational criteria), mean difference 26.3 percentage points, 95 % CI: 11.8–41.2, $p < 0.001$ (Table 4).

Participants in North America studies were less likely to have ‘moderate or better’ outcome than participants in European studies (36.4 % vs. 63.4 % for European studies, mean difference: 27.0 percentage points, 95 % CI: 4.4–49.7, $p = 0.019$). In an influence analysis, the mean estimate changed with a maximum of 10.5 percentage points. In another sensitivity analysis, a meta-regression model on ‘moderate or better’ outcome was performed, including both diagnostic and geographic groupings. The results were like the results from the single-moderator variable models, indicating independent effects of diagnostic and geographic groupings. In an influence analysis of this meta-regression model, there were no notable deviations, further lending support to the robustness of the influence of the moderator variables diagnostic and geographic groupings on the ‘moderate or better’ outcome meta-regression.

3.6. Attrition

Recovery was not associated with attrition (beta: 0.07, 95 % CI: −0.10–0.24, $p = 0.415$), but ‘good or better’ outcome was positively associated with attrition (beta: 0.49, 95 % CI: 0.01–0.98, $p = 0.048$). 'Moderate or better’ outcome had a positive but not significant association with attrition (beta: 0.37, 95 % CI: −0.15–0.89, $p = 0.166$). The general direction of the associations was that higher rates of attrition were associated with more favorable outcomes. We therefore performed a set of sensitivity analyses in which all study participants, who had dropped out of the studies, were considered having the less favorable outcome. The results for these analyses showed ‘recovery’ for 13.1 % (95 % CI: 6.4–19.7 %), ‘good or better’ for 22.4 % (95 % CI: 16.4–28.4 %), and ‘moderate or better’ for 37.9 % (95 % CI: 29.4–46.3 %) (we used the same nested groups as above).
4. Discussion

This is the first systematic review and meta-analysis of all prospective follow-up studies on schizophrenia, spanning 20 years or more. We found that 24.2% of patients with schizophrenia had 'recovered'; 35.5% had a 'good or better' outcome (which also includes 'recovered'); and 59.7% had 'moderate or better' outcome (which also includes 'good or better' and 'recovery'). This means that 40.3% had poor outcome. Though the attrition analysis indicates that the outcomes might be worse than what these figures suggest, we nevertheless conclude that a deteriorating course of illness, poor outcome, and inability to recover are not defining features of schizophrenia. While our results show a mix of outcomes, it is noteworthy that outcome of schizophrenia generally is worse than that of other psychotic disorders, e.g., psychotic mood disorders (Harrison et al., 2001; Kotov et al., 2017), substance-induced psychosis (Harrison et al., 2001; Kotov et al., 2017), or other psychosis (Harrison et al., 2001; Harrow and Joe, 2010; Kotov et al., 2017; Marneres et al., 1989; Opjordsmoen, 1991).

The heterogeneity in outcomes of schizophrenia between studies and within samples is echoed in the history of follow-up studies (Carpenter and Kirkpatrick, 1988; Häfner, 2014; Hegarty et al., 1994; Heilbrunner et al., 2016; Joe and Harrow, 2005). Our results attest to the overarching heterogeneity in 20-year outcomes of schizophrenia, with a substantial range in all outcome categories. We also found considerable methodological heterogeneity in the studies' use of scales, definitions, and cutoffs for outcome. Although some efforts have been made to make outcome definitions, measurements, and diagnostic methods more homogenous, the continuously found heterogeneity in outcomes of schizophrenia between studies inevitably raises the question to what extent it may be a product of such methodological differences. To address the issue of heterogeneity, we imposed our unified template of outcome definitions onto the studies' results, making the meta-analysis with meta-regression conceptually and methodologically stronger.

Our results on 'good or better' outcome (35.5%) are comparable to the results from the meta-analysis by Hegarty et al. (1994), who found a mean favorable outcome for 40.2%. We found a higher proportion of recovered patients (24.2%) than what was found by Jääskeläinen et al. (2013) in their meta-analysis on recovery in schizophrenia (16.4%). These two studies included long-, medium-, and short-term follow-up studies. This may indicate that after 20 years or more, patients with schizophrenia are similarly or slightly better off compared to patients with schizophrenia at earlier follow-up stages. In this context, it is relevant to mention that several studies have suggested a possible stabilization of the disorder into mild/moderate/severe outcome after 5–10 years (Davidson and McGlashan, 1997; Joe and Harrow, 2005; Ogawa et al., 1987; Wiersma et al., 2000). Ogawa et al. (1987) describe it as a scissor-phenomenon, where schizophrenia, after a tumultuous initial phase, stabilizes in either a favorable or less favorable illness trajectory. However, several studies highlight that in this stabilization phase, significant fluctuations may still occur, e.g., late-stage recovery, which suggests that there might still be room for optimism despite years of less favorable outcome (Harrison et al., 2001; Thara, 2004).

Another study that deserves mentioning in this context is The World Health Organization coordinated International Study of Schizophrenia (ISoS) study, which was excluded from our analysis as it did not meet the 20-year follow-up criterion. The ISoS study was created in an effort to reduce heterogeneity and enable cross-cultural comparison of course and long-term outcome of schizophrenia over 15–25 years (Harrison et al., 2001; Sartorius et al., 1996). They found favorable outcome for over half of the people with schizophrenia, and 'recovery' for 37.8% using M. Bleuler's rating of recovery and a GAF-disability score >60. Nonetheless, the authors also found striking heterogeneity between the participating centers, indicating that some of the differences in outcome lie beyond methodological differences and most likely pertain to the disorder of schizophrenia itself. A substantial contributor to heterogeneity in the ISoS study is what they called the 'developing country effect', referring to the frequent finding of better outcome in low- and middle-income countries (Harrison et al., 2001; Menezes et al., 2006; Perugiannis et al., 2020; Warner, 2009). Since our review only included one study from a low/middle-income country (Thara, 2004), we cannot add anything to this observation.

To further examine the heterogeneity of outcomes among the studies included in our study, we explored several possible explanations but found little that could explain it. For example, the heterogeneity could be due to change in outcome over time, but our results suggest only little change in outcome from the first long-term study enrolled patients in the early 1900ies (Ciompi, 1980) to the latest study that was initiated in 1990 (Kotov et al., 2017). This finding is somewhat consistent with that of prior meta-analyses, which found an increase in favorable outcome around and after 1950, followed by a decrease in the latest decades (1985–2011) (Hegarty et al., 1994; Jääskeläinen et al., 2013). Most likely, the increase in favorable outcome is related to the appearance of the antipsychotic drug Chlorpromazine in the early 1950's (Ban, 2007). However, Hegarty et al. (1994) suggest that while the initial increase in good outcome may reflect the improved treatment and a broadened concept of schizophrenia, the later decline in good outcome to levels like that in 1895–1955, might reflect the re-emergence of a narrower diagnostic concept.

Interestingly, we found that studies using operational diagnostic methodology (i.e., diagnostic systems from DSM-III and subsequent versions, and ICD-10) were significantly less likely to report 'good or better' (23.7% vs. 44.0%) and 'moderate or better' (43.0% vs. 69.3%) outcome than studies using a pre-operational diagnostic methodology. As suggested by Hegarty et al. (1994), our finding could be a reflection of the changing nosologically boundaries of the schizophrenia diagnosis over time. The operational definitions of schizophrenia put a stronger emphasis on the severity of psychotic symptoms (e.g., by excluding milder forms of formal thought disorders). In doing so, they favor a more chronic, delusional-hallucinatory syndrome, impeding a substantial number of hebephrenic and other non-paranoid patients, which would have received a schizophrenia diagnosis based on pre-operational definitions, to be diagnosed with schizophrenia (Jansson and Parnas, 2007; Parnas and Kendler, 2012).

Finally, we also found that studies conducted in North American were less likely to report 'moderate or better' outcome than studies conducted in European (36.4% vs. 63.4%), and, moreover, that the effects of diagnostic methodological and geographical location were independent of each other. However, since only two studies out of the four conducted in North America, could be included in this analysis, caution is warranted when interpreting the effect of this geographical location.

Despite major changes in treatment and support to people with schizophrenia over the last century, e.g., deinstitutionalization, anti-psychotic medications, psychosocial interventions, the proportion of patients with a favorable outcome appears fairly constant. However, Hegarty et al. (1994) suggest that while the initial increase in good outcome may reflect the improved treatment and a broadened concept of schizophrenia, the later decline in good outcome to levels like that in 1895–1955, might reflect the re-emergence of a narrower diagnostic concept.

4.1. Strength and limitations

Our study is strengthened by its inclusion of all prospective follow-up studies of schizophrenia of at least 20 years, and by addressing the studies' various methodological issues (outcome definitions and diagnostic criteria) and geographical location. Another strength is the present study's exclusion of studies, which grouped schizophrenia and other psychotic disorders in the same category, thereby risking overestimation of good outcome.

A limitation to this study is the heterogeneity of the reviewed studies, e.g., in terms of sample type (ranging from first-episode schizophrenia to chronic treatment refractory schizophrenia), diagnostic methods, and outcome definitions. Due to a general lack of detail in the included
studies, our review could not assess associations between treatment and outcome. This is an important limitation, since treatment likely affects outcome. Exploring the effect of treatment on outcome is an obvious target of future studies on long-term outcome of schizophrenia as well as availability of different treatment options, including treatment targeted social remission and occupational opportunities. Another limitation is the lack of information on the course of schizophrenia, which also could have been relevant to consider. Regarding our meta-analysis and meta-regression analysis, several limitations must be mentioned: Selection bias at baseline, and differences in the type of patients enrolled depending on enrollment year and region. In the meta-regression analyses, there is a general lack of power due to small sample size. The meta-analysis of recovery could only include four studies, which calls for future follow-up studies on recovery, including both clinical and social domains as well as a defined time criterion. Finally, exploring results from newer long-term studies of schizophrenia, based on the early intervention programs that were initiated in the 2000's (once these studies are published), will be highly relevant to calibrate the outcome of schizophrenia in the 21st century.

With 20+ years of follow-up there will almost inevitably be some lost to follow-up. Especially Ciompi (1980) had high attrition due to the nature of the study examining late phase schizophrenia. Our meta-regression analysis found a statistically significant association between attrition and ‘good or better’ outcome, indicating a positive bias in high attrition, and that those lost may have had a worse outcome than those included at follow-up. Our sensitivity analyses on attrition presents a ‘worst-case-scenario’ as it were, where those lost to follow-up were computed as having the less favorable outcome (which is not necessarily the case). This attrition analysis found recovery for 13.1 %, ‘good or better’ outcome for 22.4 % and ‘moderate or better’ outcome for 37.9 %. The true values are probably somewhere in between these results and those from the main analysis.

In conclusion, schizophrenia is a disorder with heterogeneous outcome. Recovery occurred in 24.2 % of patients with schizophrenia. ‘Good or better’ outcome (including also recovery) occurred in 35.5 % of patient, whereas poor outcome occurred in 40.3 % of patients. Though our attrition analyses showed that these results might overestimate favorable outcome a bit, it does not invalidate the conclusion that a non-negligible proportion of patients with schizophrenia have a more favorable outcome than what is often assumed. Still, patients with schizophrenia generally have a worse prognosis than patients with other psychotic disorders and the results from our study emphasize the need for continuous supportive care and treatment of this complex mental disorder.

Data sharing

All data included were derived from publicly available documents cited in the references. Extracted data are available upon request to the corresponding author.

Role of funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Funding

The Mental Health Services of the Capital Region of Denmark.

CRediT authorship contribution statement

I.M., J.N., R.H., A.U.P and M.G.H. designed the search strategy and selection criteria. I.M. conducted the search and removed duplicates and publications that were not peer-reviewed journal articles and not written in English. I.M. and R.H. made a full-text assessment of the remaining articles for eligibility and reviewed them independently. I.M. constructed the tables and figs. I.M., M.G.H. and J.N. wrote the first draft, which was revised by A.U.P., R.H., and J.B. J.B. conducted the meta-analysis and meta-regression, and J.B. wrote the drafts on the statistical methods and the results from the meta-analysis. All authors contributed to the revision of subsequent drafts and all authors approved final draft.

Declarations of competing interest

All authors report no conflicts of interest.

Acknowledgement

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.schres.2022.11.010.

References

Ajnakina, O., Stubbs, B., Francis, E., Goughan, F., David, A.S., Murray, R.M., Lally, J., 2020. Hospitalisation and length of hospital stay following first-episode psychotic systemic review and meta-analysis of longitudinal studies. Psychol. Med. 50, 991–1001. https://doi.org/10.1017/s0033291719000904.
Ajnakina, O., Stubbs, B., Francis, E., Goughan, F., David, A.S., Murray, R.M., Lally, J., 2021. Employment and relationship outcomes in first-episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr. Res. 231, 122–133. https://doi.org/10.1016/j.schres.2021.03.013.
American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders 4th edn, 4th edn. APA, Washington, DC.
Andreasen, N.C., 1982. Negative symptoms in schizophrenia: definition and reliability. Arch. Gen. Psychiatry 39, 784–788. https://doi.org/10.1001/
archpsyc.1982.04290070020005.
Andreasen, N.C., Olsen, S., 1982. Negative v positive schizophrenia: definition and validation. Arch. Gen. Psychiatry 39, 789–794. https://doi.org/10.1001/
archpsyc.1982.04290070020006.
Andreasen, N.C., Carpenter, W.T., Kane, J.M., Lasser, R.A., Marder, S.R., Weinerberger, D. R., 2005. Remission in schizophrenia: proposed criteria and rationale for consensus. Am. J. Psychiatry 162, 441–449. https://doi.org/10.1176/appi.ajp.162.3.441.
Angel, J., 1988. European long-term followup studies of schizophrenia. Schizophr. Bull. 14, 501-513. https://doi.org/10.1093/schbul/14.4.501.
Ban, T.A., 2007. Fifty years chlorpromazine: a historical perspective. Neuropsychiatr. Dis. Treat. 3, 495–500.
Bleuler, M., 1978. trans. by. In: Clemens, S.M. (Ed.), The Schizophrenia Disorders. Long-Term Patient and Family Studies. Yale University Press, New Haven and London.
Boonstra, N., Klaassen, R., Sytema, S., Marshall, M., De Haan, L., Wunderlin, L., Wiersma, D., 2012. Duration of untreated psychosis and negative symptoms—a systematic review and meta-analysis of individual patient data. Schizophr. Res. 142, 12–19. https://doi.org/10.1016/j.schres.2012.08.017.
Borenstein, M. (Ed.)., 2009. Introduction to Meta-Analysis. John Wiley & Sons, Chichester, U.K.
Carpenter Jr., W.T., Kirkpatrick, B., 1988. The heterogeneity of the long-term course of schizophrenia. Psychol. Bull. 104, 476–490. https://doi.org/10.1037/0033-2909.144.4.645.
Cechnicki, A., Bielanska, A., Metel, D., Susz, K., Bladzinski, P., Plencler-Banczyk, I., Kalisz, A., 2018. Comparison of the long-term treatment outcomes of women and men diagnosed with schizophrenia over a period of 20Years. A prospective study. Compr. Psychiatry 84, 62–67. https://doi.org/10.1016/j.comppsych.2018.03.011.
Cechnicki, A., Bielanska, A., Arciszewska-Lenczuk, A., Kalisz, A., Bladzinski, P., Daren, A., Metel, D., Kruk, D., 2020. Premorbid social relationship and the course of schizophrenia in a 20 year follow-up study. Psychiatr. Pol. 54, 437–451. https://doi.org/10.12740/PP/OnlineFirst/115554.
Cichocki, L., Cechnicki, A., Franczyk-Blitta, J., Bladzinski, P., Kalisz, A., Wroński, K., 2015. Quality of life in a 20-year follow-up study of people suffering from schizophrenia. Compr. Psychiatry 56, 133–140. https://doi.org/10.1016/j.
comppsych.2014.09.017.
Ciompi, L., 1980. Catamnestic long-term study on the course of life and aging of schizophrenics. Schizophr. Bull. 6, 606–618. https://doi.org/10.1093/schbul/6.4.606.
Davidson, L., McGlashan, T.H., 1997. The varied outcomes of schizophrenia. Can. J. Psychiatr. 42, 34–43. https://doi.org/10.1177/070674397904200105.
Eguma, Y., 1962. The prevention of failure in the rehabilitation of discharged schizophrenic patients (in Japanese). Psychiatr. Neurol. Jpn. 921 618. https://doi.org/10.1093/schbul/14.4.645.
Endicott, J., Spitzer, R.L., Fleiss, J.L., Cohen, J., 1976. The global assessment scale: a procedure for measuring overall severity of psychiatric disturbance. Arch. Gen. Psychiatry 33, 766–771. https://doi.org/10.1001/archpsyc.1976.017006006019.
Goldman, H.H., Skodol, A.E., Lave, T.R., 1992. Revising axis V for DSM-IV: a review of measures of social functioning. Am. J. Psychiatry 149, 1148–1156. https://doi.org/10.1176/ajp.149.9.1148.

Gross, G., Huber, G., 1986. Classification and prognosis of schizophrenic disorders in the light of the Bonn follow-up studies. Psychopathology 19, 50–59. https://doi.org/10.1159/000264429.

Grosman, L.S., Harrow, M., Rosen, C., Faul, R., Strauss, G.P., 2008. Sex differences in schizophrenia and other psychotic disorders: a 20-year longitudinal study of psychosis and recovery. Compr. Psychi. Psychiatry 49, 523–529.

Hafner, H., 2014. The Concept of Schizophrenia: From Unity to Diversity. Advances in Psychiatry Advances in Psychiatry. Article ID 929434, 39 pages.

Harding, C.M., Brooks, G.W., Ashikaga, T., Strauss, J.S., Breier, A., 1987. The Vermont. Huber, G., Gross, G., 1986. Classification and prognosis of schizophrenic disorders in that.

Kotov, R., Fochtmann, L., Li, K., Tanenberg-Karant, M., Constantino, E.A., Rubinstein, J., Jobe, T.H., Harrow, M., 2005. Long-term outcome of patients with schizophrenia: a.

Huxley, P., Krayer, A., Poole, R., Prendergast, L., Aryal, S., Warner, R., 2021. Recovery from psychotic illness: a 15- and 25-year international follow-up study. Br. J. Psychiatry 178, 506–517. https://doi.org/10.1192/bjp.178.6.506.

Harrison, G., Hopper, K., Craig, T., Laska, E., Siegel, C., Wanderling, J., Dube, K.C., 1995. schizophrenia. Schizophr. Bull. 13, 261–276. https://doi.org/10.1093/schbul/13.4.261.

Lang, F.I., Koster, M., Lang, S., Becker, T., Jager, M., 2013. Psychopathological long-term outcome of schizophrenia – a review. Acta Psychiatr. Scand. 127, 173–182. https://doi.org/10.1111/acps.12030.

Levenstein, S., Klein, D., Pollack, 1966. Follow-up study of formerly hospitalized voluntary psychiatric patients: the first two years. Am. J. Psychiatry 122. https://doi.org/10.1176/ajp.122.10.1167.

Marmers, A., Steimmeyer, E.M., Deisen, A., Rohde, A., Jüinemann, H., 1989. Long-term outcome of schizoaffective and schizophrenic disorders: a comparative study. Eur. Psychiat. Neurosci. Lett. 238, 139–139. https://doi.org/10.1007/BF00510064.

McGlashan, T.H., 1988. A selective review of recent North American long-term followup studies of schizophrenia. Schizophr. Bull. 14, 515–542. https://doi.org/10.1093/schbul/14.4.515.

McGlashan, T.H., Carpenter, W.T., 1988. Long-term followup studies of schizophrenia: editors’ introduction. Schizophr. Bull. 14, 497–500.

Menezes, N.M., Arenovich, T., Zipursky, R.B., 2006. A systematic review of longitudinal outcome studies of first-episode psychosis. Psychol. Med. 36, 1349–1362. https://doi.org/10.1017/S0033291706007951.

Newman, S.C., Bland, R.C., Thompson, A.H., 2012. Long-term course and outcome in schizophrenia: a 34-year follow-up study in Alberta, Canada. Psychiatr. Med. 42, 2137–2143.

Ogawa, K., Miya, M., Watarai, A., Nakazawa, M., Yuasa, S., Ueta, H., 1987. A long-term follow-up study of schizophrenia in Japan with special reference to the course of schizophrenia in the light of the Bonn follow-up studies. Psychopathology 19, 506–517. https://doi.org/10.1192/bjp.178.6.506.

Opjordsmoen, S., 1986. Long-term follow-up of paranoid psychoses. Psychopathology 19, 44–49. https://doi.org/10.1159/000284428.

Opjordsmoen, S., 1991. Paranoid (Delusional) disorders in the light of a long-term follow-up study. Psychopathology 24, 287–292.

Overall, J.E., Gorham, D.R., 1962. The brief psychiatric rating scale. Psychol. Rep. 10, 799–812.

Page, M.J., McKenzie, J.E., Bosuuyi, P.M., Bouton, I., Hoffmann, T.C., Mulrow, C.D., Shames, L., Tetzlaff, J.M., Akt, E.A., Brenner, S.E., Chou, R., Glanville, J., Grimsraw, J.M., Hoibrjartsson, A., Lal, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whitling, P., Moher, D., 2021. The PRISMA statement: an updated guideline for reporting systematic reviews. BMJ, n171. https://doi.org/10.1136/bmj.n171.

Parsons, J., Kendler, K., 2012. DSM-IV and the funding prototype of schizophrenia: are we regressing to a pre-Kraepelinian nosology?. In: Philosophical Issues in Psychiatry II: Nosology.

Peritogiannis, V., Gogou, A., Samakouri, M., 2020. Very long-term outcome of psychotic disorders. Int. J. Soc. Psychiatry 66, 633–641. https://doi.org/10.1177/0204040520222276.

R Core Team, 2013. R: a language and environment for statistical computing. [WWW document]. URL: http://www.R-project.org/.

Sartorius, N., Gilbainat, W., Harrison, G., Landa, E., Siegel, C., 1996. Long-term follow-up of schizophrenia in 16 countries. Soci Psychiatry Psychiatr. Epidemiol. 31, 249–258. https://doi.org/10.1007/BF00789197.

Straus, J.S., Carpenter, W.T., 1972. The prediction of outcome in schizophrenia. I. Characteristics of outcome. Arch. Gen. Psychiatry 27, 739–746.

Thara, R., 2004. Twenty-year course of schizophrenia: the Madras longitudinal study. Can. J. Psychiatry Rev. Can. Psychiatr. 49, 564–569. https://doi.org/10.1177/0706743704048080.

Velthorst, E., Fett, A.K.J., Reichenberg, A., Perlm, G., van Os, J., Bremot, E.J., Kotov, R., 2017. The 2017.20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. Am. J. Psychiatry 174, 1075–1085.

Viciethauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36, 1–48. https://doi.org/10.18637/jss.v036.i03.

Volavka, J., Vevera, J., 2018. Very long-term outcome of schizophrenia. Int. J. Clin. Pract. 72, e13094. https://doi.org/10.1111/i jcp.13094.

Warner, R., 2009. Recovery from schizophrenia and the recovery model. Cur. Opin. Psychi. 22, 374–380.

Wells, G., Shea, B., O’Connell, D., Robertson, J., Peterson, J., Losos, M., Tugwell, P., 2008. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomized Studies in Meta-Analysis.

Wierriena, D., Wanderling, J., Drongirebeca, E., Ganev, K., Herring, G., An Der Heiden, W., Nienhuis, J.F., Walsh, D., 2000. Social disability in schizophrenia: its development and prediction over 15 years in incidence cohorts in six European centres. Psychol. Med. 30, 1155–1167.

World Health Organisation (WHO), 1988. WHO Psychiatric Disability Assessment Schedule (WHO/DAS). WHO, Geneva.