Content and quality of 2000 controlled trials in schizophrenia over 50 years

Ben Thornley, Clive Adams

Abstract

Objective To provide a comprehensive survey of the content and quality of intervention studies relevant to the treatment of schizophrenia.

Design Data were extracted from 2000 trials on the Cochrane Schizophrenia Group's register.

Main outcome measures Type and date of publication, country of origin, language, size of study, treatment setting, participant group, interventions, outcomes, and quality of study.

Results Hospital based drug trials undertaken in the United States were dominant in the sample (54%). Generally, studies were short (54% < 6 weeks), small (mean number of patients 65), and poorly reported (64% had a quality score of ≤ 2 (maximum score 5)). Over 600 different interventions were studied in these trials, and 640 different rating scales were used to measure outcome.

Conclusions Half a century of studies of limited quality, duration, and clinical utility leave much scope for well planned, conducted, and reported trials. The drug regulatory authorities should stipulate that the results of both explanatory and pragmatic trials are necessary before a compound is given a licence for everyday use.

Introduction

The advent of randomised controlled trials coincided with a revolution in the care of people with schizophrenia. Drug treatments were developed that dramatically improved the mental state of those for whom little hope had previously existed.[1] Psychiatrists welcomed the randomised trial, and a tradition of evaluative research was strengthened.

The creation of registers of randomised controlled trials, such as that developed by the Cochrane Schizophrenia Group,[2] affords an opportunity to assess the quality and content of evaluative research in well defined sampling frames. There are now many examples of such surveys in specific journals,[3–5] including the pilot study for this work,[6] but research into the quality and content of trials in specific healthcare specialties is less common. We focused on the care of patients with schizophrenia or other non-affective psychoses to replicate and expand work in other specialties.[7–11]

Methods

Inclusion criteria—Every report of the first 2000 trials on the Cochrane Schizophrenia Group's register was eligible. The register contains reports of published and unpublished randomised controlled trials and controlled clinical trials (parallel group comparative studies in which allocation of treatment is not explicitly stated to be random). These studies relate to the care of those with schizophrenia and other non-affective psychoses.

Identification and selection of studies—Key journals were identified and hand searched from 1948 to December 1997. Conference proceedings were also hand searched. We comprehensively searched *Biological Abstracts* (1982-96), CINAHL (1980-96), the Cochrane Library (issue 3, 1997), Embase (1980-96), LILACS (1980-96), Psyc/Lit (1974-96), PSYNDEX (1980-95), Medline (1966-96), and Sociofile (1985-96).[12] The resulting 30 000 electronic records were checked for duplicates before we highlighted studies that were possibly relevant. We obtained 6000 full copies, which were added to the register if they met the inclusion criteria as outlined above. The final sample was 3181 publications (around 2500 trials). Time constraints forced us to survey so the first 2000 trials (reported in 2275 publications).

Extraction and analysis of data—We recorded the type and date of publication, country of origin, and language of each study. We used a measure of methodological quality based on each trial's description of randomisation, blinding, and withdrawal from treatment.[13] The maximum score was 5, for which the report had to have given appropriate methods of generating random assignment, appropriate blinding of participants and raters, and details on those who withdrew from the trial before its conclusion. This measure was chosen for its validity[13] and ease of use and because low scores, indicating poor quality of reporting, are associated with an increased estimate of benefit.[14] It does not specifically rate concealment of allocation, which is also related to trial outcome.[15] We also recorded the size of the study, treatment setting, participants, interventions, and outcomes, and all data were stored in spreadsheets. BT coded most of the reports. CA recoded a 10% sample to test and ensure reliability. Data were analysed with Microsoft Excel.

Results

Coding of variables was reliable. There was over 90% agreement in all but the numbers completing the study (70%) and listing of outcome instruments; in about 10% of reports the principal rater (BT) failed to identify one of the scales, often among several used.

Over 95% (1954) of the 2000 trials were in people with schizophrenia, serious or chronic mental illness, psychosis, or movement disorders. Most of the 2275 reports were fully published in journals (1940, 85%), while the remainder were presented at conferences (253, 11%) or published as letters, in books, as chapters in books, or as product monographs (82, 4%). The *BMJ* and *Lancet* publish a few more schizophrenia trials than *JAMA* and the *New England Journal of Medicine* (21, 33, 6, 2 respectively), but all these widely read journals were limited sources of trials on this most serious, costly illness. Most trials were published in general psychiatric journals.

Cochrane Schizophrenia Group, Summertown Pavilion, Oxford OX2 7LG
Ben Thornley, *research assistant*
Clive Adams, *coordinating editor*

Correspondence to:
Dr Adams
clive.adams@psych.ox.ac.uk
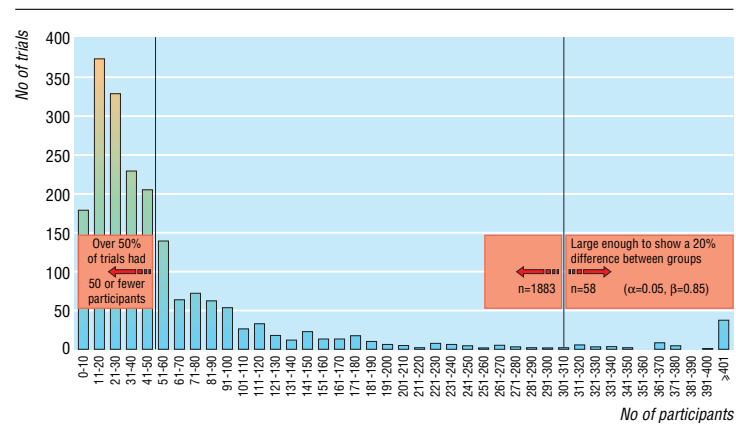
Size of trials (n=1941; 59 studies did not report study size)

The number of trials relevant to schizophrenia rose steadily with time, from about 20 per year in the 1950s and 1960s to an average of nearly 75 per year in the past decade ($\beta = 1.9$, $r^2 = 0.59$, $P < 0.001$, where β is the estimated yearly change and r^2 the yearly data explained by a linear trend).

Most (2214, 97%) of the reports were published in English. Most (1238, 54%) were from North America, with 37% (849) from Europe and 8% (188) from the rest of the world. Trial output from North America increased at a faster rate than that from Europe and the rest of the world (0.9, 0.7, and 0.3 extra trials per year respectively).

The quality of reporting was poor. Only 4% (80) of the trials clearly described the methods of allocation. Explicit descriptions of blinding were adequate in only 22% (440) of trials, while some description of treatment withdrawals was given in 42% (840). One

Table 1 Interventions investigated in 2000 controlled trials of schizophrenia treatment

Type of treatment* and most commonly used type	No (%) of trials with intervention
Drugs (n=437)	1725 (86)
Antipsychotics (n=135)	1187 (59)
Phenothiazines (n=50)	691 (35)
Butyrophenones (n=13)	339 (17)
Atypical agents (n=12)	160 (8)
Thioxanthenes (n=10)	132 (7)
Antidepressants (n=31)	127 (6)
Antidyskinetic agents (n=31)	138 (7)
Anxiolytics and hypnotics (n=31)	91 (5)
Psychotherapy (n=85)	164 (8)
Family therapy	39 (2)
Group therapy (unspecified)	24 (1)
Social skills training	12 (1)
Individual therapy (unspecified)	12 (1)
Policy or care package (n=48)	172 (9)
Standard care	109 (6)
Case management	31 (2)
Hospital care	31 (2)
Community care	14 (1)
Physical (n=21)	77 (4)
Electroconvulsive therapy	55 (3)
Insulin coma treatment	11 (1)
Haemodialysis	8 (<1)
Other (n=31)	37 (2)

*n=total number of different interventions.

per cent (20) of the 2000 trials achieved a maximum quality score of 5. Just under two thirds (1280) scored 2 or less, which means that they barely, if at all, described any attempt to reduce the potential for introduction of bias at allocation or rating of outcome, placebo effects, or the fate of all participants. A score of 3 or more was predefined as better quality. Just 33% (354/1062) of North American trials achieved this, compared with 36% (262/724) of European trials and 43% (77/180) of those from the rest of the world ($\chi^2 = 9.23$, $P < 0.01$). Studies from Canada (n = 103) and a combined group of the Middle East and Asia (n = 109) were particularly well reported (98, 46% scoring 3 or higher). We found little evidence that the quality of trial reporting improved with time. From 1950 to 1997 the mean quality score was consistently under 2.5.

The average number of trial participants was 65, with no discernible change over time ($\beta = 0.2$, $r^2 = 0.7$, $P = 0.4$). Only 20 trials (1%) raised the issue of the statistical power of the study. The average size of schizophrenia trials was small. For an outcome such as clinically important improvement in mental state to show a 20% difference between groups a study would have to have 150 participants in each arm ($\alpha = 0.05$, power 85%). Only 3% (60) of studies were of this size or greater. More than 50% of trials had 50 or fewer participants (figure).

On average, just under 12% of participants left the studies early, although the trend was towards increasing loss to follow up ($\beta = 0.01$, $r^2 = 0.6$, $P < 0.001$). Over half of the trials lasted six weeks or less (1082, 54%), and less than one fifth allowed more than six months to evaluate the treatments (382, 19%).

Only 272 (14%) of the total sample of trials were clearly community based, but the proportion increased ($\beta = 0.1$, $r^2 = 0.92$, $P < 0.01$). Even in the 1990s, however, the proportion was still small (23%, 135/587).

Interventions were classed as drug treatment, psychotherapy (any treatment based on talking), physical treatment (electroconvulsive therapy, psychosurgery), policy or care packages (case management, team treatment), and other (table 1). Overall, 1725 (86%) of the 2000 trials evaluated the effects of 437 different drugs. Haloperidol was an increasingly frequent comparator ($\beta = 0.5$, $r^2 = 0.6$, $P < 0.001$). Overall, the proportion of drug trials declined somewhat over time ($\beta = -0.002$, $r^2 = 0.5$, $P = 0.03$), with studies of psychotherapy and policy or care packages increasing.

In all, 510 (25%) studies did not use rating scales to measure outcomes. The remaining 1490 trials used 640 different instruments. These were broadly classified; table 2 lists the most popular. Overall, 369 scales were used only once. Most trials used between one and five instruments, but greater numbers were not uncommon, with one trial using 17 different outcome scales.

Discussion

Sampling biases
Our sample of 2000 trials is likely to be biased in some respects. The searching was largely, but not exclusively, in English, and our ability to code articles in languages other than English was limited. An undiscovered body of large, high quality trials published in languages

Table 2 Outcome instruments used in 2000 controlled trials of schizophrenia treatment

Outcome measured* and most commonly used instrument	No (%) of trials
Psychiatric symptoms (n=194)	1250 (63)
Brief psychiatric rating scale	800 (40)
Scale for assessment of negative symptoms	113 (6)
Inpatient multidimensional psychiatric rating scale	68 (3)
Positive and negative syndrome scale for schizophrenia	67 (3)
Cognitive functioning (n=97)	141 (7)
Wechsler adult intelligence scale	24 (1)
Digit symbol test	18 (1)
Continuous performance task	14 (1)
Wechsler memory scale	13 (1)
Behaviour (n=80)	367 (18)
Nurses observation scale for inpatient evaluation	178 (9)
Wing-Ward behaviour scale	41 (2)
MACC behavioural adjustment scale	12 (1)
Baker-Thorpe rating scale	11 (1)
Side effects (n=67)	431 (22)
Simpson-Angus scale	175 (9)
Abnormal involuntary movement scale	114 (6)
Extrapyramidal symptom rating scale	43 (2)
Treatment emergent symptoms scale	36 (2)
Social functioning (n=66)	127 (6)
Katz adjustment scales	23 (1)
Social adjustment scale	20 (1)
Global assessment of function scale	10 (1)
Evaluation of social functioning form	8 (<1)
Neurological and psychomotor functioning (n=41)	92 (5)
Reaction time tests	17 (1)
Finger tapping test	15 (1)
Handwriting tests	13 (1)
Neurological rating scale	12 (1)
Activities of daily living (n=34)	73 (4)
Katz adjustment scales	23 (1)
Hospital adjustment scale	6 (<1)
Activities of daily living	5 (<1)
Level of functioning scale	5 (<1)
Global measures (n=20)	392 (20)
Clinical global impression	331 (17)
Global assessment scale	46 (2)
Global assessment of function scale	10 (1)
Global evaluation scale	3 (<1)
Other (n=115)	188 (9)
No specific instrument used	510 (26)

MACC=motility, affect, cooperation, communication.
*n=total number of scales.

other than English is unlikely as researchers everywhere are conscious that the common language of the scientific community is English. The first 2000 trials on the Cochrane Schizophrenia Group's register were a subsample of around 2500 eventually identified by the search. High availability of a report would probably have increased the chance of early entry on to the register and so of being within the survey sample. However, this potential selection bias would have been offset by our limited efficiency. The register was compiled over two years, so we had a mixture of easily acquired reports and those that had been difficult to find. The study sample, however, may well incorporate less selection biases than the trials readily available through Medline and is representative of those on the Cochrane controlled trials register, the most comprehensive source of randomised controlled trials and controlled clinical trials.[16]

The profile of trials

Most schizophrenia trials are undertaken in North America. The United States, in particular, has a strong tradition of evaluative research in randomised controlled trials, but only 2% of the world's population of people with schizophrenia live in North America. How applicable the findings of these trials are to the 43 million other patients in Africa, Australasia, and Europe is difficult to assess. Further problems with generalisability arise from the fact that most participants in trials were people in hospital, even in the 1990s.

The quality of reporting in this large sample of trials was poor and showed no sign of improvement over time. As low quality scores are associated with an increased estimate of benefit,[14] schizophrenia trials may well have consistently overestimated the effects of experimental interventions. We hope that this will change with wider adoption of the CONSORT recommendations.[17] However, although quality of reporting has been a proxy measure of methodological quality of a trial,[14 15 18] cosmetic adherence to CONSORT requirements might mask low grade studies.

Drug treatments are the bulwark of treatment of schizophrenia, so it is not surprising that drug trials dominate the sample. Most important drug trials in recent years (mostly of the new atypical generation of antipsychotic drugs, such as risperidone and olanzapine) use haloperidol as the control. This drug is likely to give obvious side effects that render successful blinding difficult, if not impossible, probably making the outcomes used more vulnerable to bias. In addition, because haloperidol is also a potent cause of adverse effects, most drugs to which it is compared will have favourable side effect profiles. Therefore, so long as the new experimental drug has moderate antipsychotic properties, favourable outcomes can be expected. Comparisons with other old, but less toxic, antipsychotic drugs, such as medium doses of thioridazine or sulpiride, are rare.

The lack of statistical power was reflected in the use of an extraordinary number of rating scales. It is often possible to achieve significance on these fine measures with small numbers. These devices by researchers leave unaddressed the clinical interpretation of these measures and the fact that scales are rarely used in clinical practice. To complicate matters further scales, and subscales, were often used at frequent intervals within the trial. The sheer quantity of data testing will then result in misleading, significant findings appearing by chance. The use of scales in schizophrenia trials is the focus of ongoing work. Further difficulties with using the evidence generated by this mass of research are that the studies are of limited duration for an illness that often lasts decades.

Room for improvement

The findings of this survey are as bad, if not worse, as those for other disciplines of health care.[8–11 14 15] Certainly, there is a long way to go before all interventions for patients with schizophrenia have been adequately evaluated and systematically reviewed and some of the enduring questions about the efficacy of treatment are answered. There is great scope for well conceived, conducted, and reported trials. The use of haloperidol as a control and rating scales of little clinical utility may well be fostered by the stipulations of

Key messages

- The advent of randomised controlled trials coincided with many new drug treatments for schizophrenia

- This survey of 2000 randomised controlled trials of treatment for schizophrenia found that the reporting of key aspects of trial methods could easily be improved

- The consistently poor quality of reporting is likely to have resulted in an overoptimistic estimation of the effects of treatments

- Large studies, of long duration, investigating outcomes of importance to clinicians and patients are needed

various drug regulatory bodies. It should not be beyond the ability of a well motivated pharmaceutical industry to negotiate with these bodies to ensure that both explanatory and pragmatic studies[19] are required for licensing.

1 Freeman H. The tranquilizing drugs. In: Bellek L, ed. *Schizophrenia: a review of the syndrome.* New York: Logos, 1958:473-500.
2 Adams CE, Duggan L, Whalbeck K, White P. The Cochrane Schizophrenia Group. *Schizophr Res* (in press).
3 Fahey T, Hyde C, Milne R, Thorogood M. The type and quality of randomized controlled trials (RCTs) published in UK public health journals. *J Public Health Med* 1995;17:469-74.
4 Lent V, Langenbach A. A retrospective quality analysis of 102 randomized trials in four leading urological journals from 1984-1989. *Urol Res* 1996;24:119-22.
5 Silagy CA. Developing a register of randomised controlled trials in primary care. *BMJ* 1993;306:897-900.
6 Ahmed I, Soares KVS, Seifas R, Adams CE. Randomized controlled trials in Archives of General Psychiatry (1959-1995): a prevalence study. *Arch Gen Psychiatry* 1998;55:754-5.
7 Chalmers I. A register of controlled trials in perinatal medicine. *WHO Chron* 1986;40:61-5.
8 Chalmers I, Hetherington J, Newdick M, Mutch L, Grant A, Enkin M, et al. The Oxford Database of Perinatal Trials: developing a register of published reports of controlled trials. *Control Clin Trials* 1986;7:306-24.
9 Cheng K, Ashby D, O'Hea U, Smyth R. The epidemiology of randomised controlled trials in cystic fibrosis. *Israel J Med Sci* 1996;32:S260.
10 Nicolucci A, Grilli R, Alexanian AA, Apolone G, Torri V, Liberati A. Quality, evolution, and clinical implications of randomized, controlled trials on the treatment of lung cancer. A lost opportunity for meta-analysis. *JAMA* 1989;262:2101-7.
11 Vandekerckhove P, O'Donovan PA, Lilford RJ, Harada TW. Infertility treatment: from cookery to science. The epidemiology of randomised controlled trials. *Br J Obstet Gynaecol* 1993;100:1005-36.
12 Adams CE, Duggan L, Roberts L, Wahlbeck K, White P. The schizophrenia module. *Cochrane Library* [database on disk and CD ROM]. Cochrane Collaboration; 1997, Issue 1. Oxford: Update Software, 1997. Updated quarterly.
13 Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
14 Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
15 Schulz KF, Chalmers I, Altman DG, Grimes DA, Dore CJ. The methodologic quality of randomization as assessed from reports of trials in specialist and general medical journals. *Online J Curr Clin Trials* 1995;Doc No 197:[81 paragraphs].
16 Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-9.
17 Chalmers TC, Celano P, Sacks HS, Smith H Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983;309:1358-61.
18 Cochrane Controlled Trials Register. In: *The Cochrane Library* [database on disk and CD ROM]. Cochrane Collaboration. Oxford: Update Software, 1996. (Updated quarterly.)
19 Roland M, Torgerson DJ. What are pragmatic trials? *BMJ* 1998;316:285.

One hundred years ago
Labour and refreshment

Hard-working men, especially in the learned professions are often puzzled as to the right hours for their meals and the right diet to be taken at their meals. The much graver question of stimulants also interests them personally. The fear of a mid-day meal is very common amongst brain-workers. A good luncheon is apt to interfere with inspiration, especially if any form of alcohol be taken. Tea, very popular with scholars, is less objectionable as to its immediate effects, but it is very dangerous to digestion, especially amongst scholars who, like workwomen, nearly always take too much and prefer their tea to be strong, "red as blood," as Garrick used to say. The chief danger in taking too light a mid-day meal is that the student is at his worst, physically, just before dinner time. Hence dinner is rendered fresh source of dyspepsia, and if he strolls about before the evening meal he is apt to take cold. The special stimulants of great men have been recorded in their memoirs. One of most remarkable records of this kind has come to light in Aubrey's *Brief Lives*, founded on notes taken by the author in the middle of the seventeenth century. Few drinks are more distrusted amongst brain-workers in these days than ale, but Prynne seems to have thought otherwise. He was a voluminous writer, and his pamphlets lost him his ears and endangered his head. Aubrey informs us that "About every three hours his (Prynne's) man was to bring him a roll and pot of ale to refocillate his wasted spirits; so he studied and drank and munched some bread, and this maintained him till night; and then he made a good supper. Now he did well not to dine which breaks off one's fancy, which will not presently be regained." The danger of *a priori* reasoning may be seen from the above quotation. Those who have forgotten about Prynne would take him from this note of his diet to be a Bohemian of a well-known old-fashioned type; yet he was a stern uncompromising Puritan who suffered mutilation and the pillory, and had a violent and very un-Bohemian prejudice against the stage. The diet does not then proclaim the man. We doubt, however, if the serious student of the end of the nineteenth century would do well to "refocillate his wasted spirits" by drinking a pot of ale every three hours. Prynne was perhaps wise not to dine, which meant to lunch in those days, and implied a very heavy meal. But the munching of rolls all day and the "good supper" were questionable. His style was very caustic; perhaps because of the misrule of Charles I, but possibly also through dyspepsia, which also had so disastrous an influence on the literary temper of Thomas Carlyle.
(*BMJ* 1898;ii:1088)