

Measurement of long-term outcomes in observational and randomised controlled trials

RICHARD HODGSON, CHRIS BUSHE and ROBERT HUNTER

Background Randomised controlled trials (RCTs) are the gold standard for evaluating treatment efficacy. However, the outcomes of RCTs often lack clinical utility and usually do not address real-world effectiveness.

Aims To review how traditional RCTs may be triangulated with other methodologies such as observational studies and pragmatic trials by highlighting recently reported studies, outcomes used and their respective merits.

Method Literature review focusing on drug treatment.

Results Recently reported observational and some pragmatic studies show a degree of consistency in reported results and use outcomes that have face validity for clinicians.

Conclusions No single experimental paradigm or outcome provides the necessary data to optimise treatment of mental illness in the clinical setting.

Declaration of interest R.H. and R.H. have received funding from several pharmaceutical companies. C.B. is an employee of Eli Lilly UK. Funding detailed in Acknowledgements.

Evaluating treatment outcomes in mental illness presents unique and formidable challenges. The natural course of many psychiatric disorders is cyclical with spontaneous remission a distinct possibility (Ciompi, 1980). Environmental factors are important but poorly understood. Mental illness continues to be characterised in terms of symptoms despite advances in understanding pathogenesis. Currently, most published pharmacotherapy clinical trial data derive from trials performed to prove efficacy and safety to regulatory authorities. Thus clinicians making treatment decisions are commonly presented with a series of randomised controlled trials (RCTs) undertaken to meet regulatory requirements, with outcomes that are neither pragmatic nor easily transferable to clinical practice.

It is assumed that psychiatrists will base their treatment on the best available evidence but what is the best available evidence for a given clinician? Many factors are relevant and include personal experience, the literature, anecdote, opinion leaders, the pharmaceutical industry, guidelines and cost. However, little is known about actual prescribing and other treatment decisions (Hoblyn *et al*, 2006). Clinicians, purchasers and user advocates are also demanding more pragmatic end-points, and longer trials have shown the utility of relapse rates, hospitalisation and discharge rates as outcome measures (Csernansky *et al*, 2002).

Thus in 2007 'best available evidence' is generally accepted as the RCT, but the available RCT evidence is at best incomplete, and at worst, flawed (Black, 1996). The aim of this paper is to show practising clinicians the spectrum of quantitative evidence and pragmatic outcomes.

EVOLUTION OF CLINICAL TRIALS

Since the 1940s the RCT has been the principal method of comparing the efficacy of

all forms of medical treatment, and the basic concept has been developed and refined to further reduce bias. This has been evident in psychiatry with the development of rating scales and classification systems which enhance reliability, if not always validity. The RCT has informed the development of evidence-based medicine, meta-analysis and the Cochrane Collaboration. Evidence-based medicine resulted in part from the realisation that clinical practice is often poorly informed by the best available evidence, and that many widely used treatments are either untested or have been shown to be ineffective (Lenzer, 2004). Evidence-based medicine has also been seen as a means by which policy makers, sometimes with academic support, control clinical freedom (Williams & Garner, 2002). Although RCTs have resulted in the discontinuation of fashionable but ineffective treatments such as insulin coma therapy (Ackner & Oldham, 1960), they are not without problems (Thornley & Adams, 1998). More recently other paradigms, including observational and pragmatic studies (Roland & Torgerson, 1998), have gained in acceptance and been recommended as having a useful role in evaluation of treatment by the National Institute for Health and Clinical Excellence (National Institute for Clinical Excellence, 2002).

RANDOMISED CONTROLLED TRIALS

In general an RCT assesses efficacy – whether the treatment works in a controlled environment – not whether it works in the real world (effectiveness) (Table 1). Many factors affect the relationship between efficacy and effectiveness. This is acknowledged in the CONSORT criteria for RCTs by the need to assess the generalisability of the results, although a framework for assessing and reporting this is lacking (Bonell *et al*, 2006). Trials have been criticised for not adhering to CONSORT guidelines, but even apparent adherence can lead to challenges (El-Sayeh *et al*, 2006).

Patient recruitment and selection bias

Whether clinically significant selection bias occurs during recruitment to clinical trials is contentious. Although Burns (2006) reported that the basic demography of patients in a large naturalistic study was

Table 1 Comparison of key features of randomised controlled trials and observational studies

Randomised controlled trial	Observational study
Modest numbers of patients	Large number of patients
Modest duration	Longer duration
High drop-out rate	Lower drop-out rate
Statistically significant results	Clinically meaningful results
Structured dosing regimen	Naturalistically selected dosing
Randomisation	Naturalistic treatment selection
Maximises internal validity	Maximises external validity
Minimal bias and variability	Generalisability
Homogeneous patient population	Heterogeneous patient population
Artificial adherence and population	Adherence not mandated, 'real' patients
Demonstrates efficacy	Assesses effectiveness
Excludes confounding treatments	Concomitant treatments allowed
Complex applied scales	Outcomes used in everyday clinical practice
Outcomes generally symptom focused	Outcomes include cost, adherence, resource use

similar to that of a widely reported RCT, other authors have noted that the more chaotic patient who is difficult to manage will not be entered into a clinical trial as, even if they consent, they will undoubtedly drop out of follow-up (Lester & Wilson, 1999; Harrison-Read *et al*, 2002). Trials rarely report the number of patients considered or screened for a trial who are never included. Although this is a CONSORT requirement, clinicians will make pre-screening decisions regarding eligibility that are never reported. This is a potential source of bias and might limit extrapolation of results. It is likely that these difficulties are a serious unreported bias in published RCTs for psychological treatments. For example, reviews of the impact of day hospital treatment have failed to take entry criteria into account, leading to potentially erroneous conclusions (Thornicroft & Strathdee, 1994). The need for informed consent might inadvertently affect the generalisability of data from RCTs. All trials of intramuscular olanzapine (Meehan *et al*, 2001; Wright *et al*, 2001) were conducted in patients who gave informed consent and, although positive, the results cannot be interpreted as indicating that the drug will be as effective in patients who are highly disturbed.

Although biases are reduced in RCTs they are not eliminated, and indeed specific biases may even be created. Aside from the increased practical difficulties of including older adults in clinical trials, only 4.2% of older patients with major depression meet the increasingly rigorous inclusion and exclusion criteria of phase 3 studies

(Yastrubetskaya *et al*, 1997). Women have sometimes been underrepresented in RCTs primarily because of concerns regarding conception while on trial medication, although this may be changing.

Patients with comorbid disorders are usually excluded from RCTs and this does not allow trials to reflect the rate of substance misuse and physical ill health in people with mental illness (Phelan *et al*, 2001). Previous exposure to trial medication is often unreported, but McQuade *et al* (2004) reported that 25% of patients in this randomised trial had prior exposure to one of the evaluated drugs. Generally, RCTs do not control for previous number of admissions or other markers of 'difficult to treat' patients (Hodgson *et al*, 2005). This might lead to newer treatments being tried in patients who are more difficult to treat, which may lead to suboptimal results for newer treatments (Davis *et al*, 2003).

Rating scale outcomes

The outcome measures used in RCTs affect the generalisability of the results. Although these outcome measures have been refined over decades to improve reliability, in studies their use may affect the face validity of the results. Clinicians would have difficulties in understanding what a fall of 20% in score on the Positive and Negative Syndromes Scale (PANSS; von Knorring & Lindstrom, 1995) means in clinical practice. Indeed Kane *et al* (1988) suggested this as an outcome only for treatment-resistant patients and a recent analysis (Leucht *et al*, 2005) has shown that a drop of 50%

in PANSS score may better equate to a Clinical Global Impression Scale (CGI; Haro *et al*, 2003) rating of 'much improved'.

Pragmatic outcomes

Rating scales might not reflect clinical reality and there may be dissonance between rating scale response and a pragmatic clinical end-point such as discharge from hospital (McCue *et al*, 2006). Pragmatic research and outcomes focus on whether an intervention works under real-life conditions and whether it works in terms that matter to the patient. However, if broader concepts are used, such as remission, relapse or rehospitalisation, then other problems emerge. Rehospitalisation is easily measured, but in an individual trial may be mediated by other variables such as admission criteria. Remission or response rates might have more clinical utility but have been criticised on the grounds of variability of results if an arbitrary cut-off is used, although sensitivity analysis can be used to assess the effect of changing parameters (Linden *et al*, 2006; van Os *et al*, 2006).

Rates of discontinuation of treatment may be a proxy for treatment effectiveness (Hodgson, 2005; Lieberman *et al*, 2005; Kinon *et al*, 2006). Kinon *et al* (2006) undertook a meta-analysis of RCTs of atypical antipsychotics using reported discontinuation as an outcome and found far more variability between drugs than might have been anticipated from the headline results, which usually (marginally) favour the sponsor's product (Heres *et al*, 2006). Further exploration of these pragmatic end-points in long-term studies facilitate a better understanding of the face and predictive validity of rating scales. Any dissonance between comparator drugs using varied end-points might be cause for concern. A recent non-inferiority RCT comparing two atypical antipsychotics at 1 year showed consistency of superiority for one in parameters ranging from PANSS score to discontinuation and hospitalisation rates (www.clinicalstudyresults.org/drugdetails/?drug_name_id=187&sort-c.company_name&page=1&drug_id=509). However, use of outcomes such as hospitalisation might preclude cross-service comparisons. Quality of life has also been used as an outcome but although such measures are laudable, in practice the outcomes are difficult to measure and may not be amenable to change (Boardman *et al*, 1999).

Tolerability

Published RCTs have been criticised for inadequate reporting of side-effects and adverse events (Ioannidis & Lau, 2001; Papanikolaou *et al*, 2004). The incidence is usually reported but duration and severity are not. These are important variables and may make the difference between persevering with medication or abandoning a therapeutic trial. For data such as prolactin levels RCTs often report mean cohort values rather than pragmatically useful categorical rates (Bushe & Shaw, 2007).

Study length and drop out

Typically patients in secondary services receive treatment for periods of time that far exceed those of RCTs, which are often as short as 4 weeks. The Schizophrenia Outpatient Health Outcomes (SOHO) study (Haro *et al*, 2006) demonstrated continued improvement over 3 years. Short RCTs will not assess all tolerability issues and whether improvement is maintained. However, RCTs are getting longer (Lieberman *et al*, 2003; McQuade *et al*, 2004). The corollary of longer study periods is lower follow-up rates and, paradoxically, high follow-up rates might be an indicator of a biased study population. Drop-out rates over 6 weeks are on average 35% and at 6 months can be around 72% (Leucht *et al*, 2003; McQuade *et al*, 2004), making interpretation of data complex.

Randomised controlled trials are designed to minimise bias and in creating this artificial environment treatment effects may be obviated. Although the true masking of many trials has been debated (Moncrieff, 1997), clinicians cannot intervene in trials in a timely or appropriate manner. Doses and visits are predetermined, as is the ability to respond to potential side-effects. These issues are relevant to the placebo arm, as often placebo group patients are receiving a psychoactive drug such as lorazepam (Meehan *et al*, 2001; Wright *et al*, 2001). Randomised controlled trials are often designed to fulfil regulatory requirements to obtain marketing authorisations for a new drug. There will be significant delays between study conception, recruitment, follow-up and publication of results. Clinicians often anticipate this with off-label prescribing (Hodgson & Belgamwar, 2006). The reality is that few RCTs are ever undertaken by pharmaceutical companies after launch. This is for many reasons, including the

relatively short patent life. Thus, when such RCTs are performed there is often a perceived need for the data to be available quickly. Rarely are these trials long term.

Evolution of the RCT paradigm is seen in the CATIE trial (Lieberman *et al*, 2005; Table 2). In addition to traditional outcome measures, continuation on an antipsychotic was used as an outcome. Such an outcome should resonate with clinicians as medication is most commonly discontinued owing to lack of effectiveness or side-effects (Hodgson, 2005). Meta-analysis shows that lack of effectiveness is the major reason for discontinuation and differentiates between atypical antipsychotics in RCTs. In contrast, discontinuation for side-effects is relatively uniform (Kinin *et al*, 2006).

For the reasons above, RCTs fail to provide the clinician with all the necessary information to prescribe confidently. In order to prescribe a new product the clinician uses previous experience, critical review of early results and the experience of others. In other words the clinician is in effect, albeit informally, undertaking a naturalistic/observational study. The definition of an observational study can be problematic, but in the context of this paper we have identified the key element as a research design where the allocation of treatment is not fully under the control of the researcher (Table 1).

OBSERVATIONAL STUDIES

Limitations

There are notable long-term observational follow-up studies in psychiatry (Ciompi, 1980; Harding, 1988) which illustrate the natural history of schizophrenia over decades. Given this expertise, it is perhaps surprising that there are so few studies looking at treatment effects over the longer term, especially as many potential outcome measures could be collected routinely. Observational studies have design faults that limit their interpretation (Table 1). Most importantly, true randomisation cannot occur in an observational study. However, the strengths of observational studies mirror the weaknesses of RCTs, and it is for this reason that National Institute for Health and Clinical Excellence (NICE) has argued for well-conducted observational studies to demonstrate effectiveness. Observational studies might also represent the only method for studying certain aspects of treatment when masking is not possible or ethical concerns preclude randomisation (Cook &

Campbell, 1979). Indeed, in service evaluation studies randomisation may interfere with the dependent variable and observational studies often exploit service inequalities (Dean *et al*, 1993). Another potential bias in observational studies is rating bias, although the SOHO study has shown high correlations between clinician and patient ratings. With end-points such as hospitalisation, bias is minimised, especially if these data are collected routinely (Hodgson *et al*, 2001).

Observational studies have been criticised because they are believed to overestimate treatment effects. However, recent comparison between RCTs and observational studies does not support this view (Benson & Hartz, 2000; Concato *et al*, 2000; Kasper *et al*, 2001). Concato *et al* (2000) challenge the accepted hierarchy of clinical designs by reviewing outcomes from various methodologies in a variety of study areas and conclude that observational studies neither over- nor underestimate treatment effects to any significant degree. They opine that observational studies are more likely to produce homogeneous results as they include a broad spectrum of the population at risk. In addition, there is less chance of systematic treatment biases because of the broad treatment population.

Recent observational studies

The CATIE study (Lieberman *et al*, 2005), an RCT sponsored by the National Institute of Mental Health, compared the outcome of atypical antipsychotics with the typical antipsychotic perphenazine and also incorporated a switching strategy to evaluate clozapine. The results mirror those of Tiihonen *et al* (2006) in that clozapine and olanzapine were the only oral atypical antipsychotics to demonstrate lower discontinuation rates when compared with oral first-generation and other second-generation antipsychotics. The study reported by Tiihonen *et al* (2006) is particularly noteworthy as it follows a nationwide cohort of over 2000 people with first-episode schizophrenia for up to 7 years. In addition to showing differences in rehospitalisation and relapse rates between commonly available antipsychotics in Finland, it also shows the effectiveness of medication in reducing suicide and physical morbidity (adjusted relative risk 37.4, 95% CI 5.1–276 and 12.3, 95% CI 6.0–24.1 respectively). The relative therapeutic effects of the drugs studied did not vary

Table 2 Key recent observational and pragmatic studies and randomised controlled trials in schizophrenia

Reference	Methodology	Study size and follow-up	Setting	Key outcome measures	Key findings	Funding source
Hodgson <i>et al</i> (2005)	Observational	502 patients up to 7 years	England	Medication discontinuation	Lowest discontinuation rate with clozapine, then olanzapine, then risperidone	Unrestricted grant from pharmaceutical industry
Haro <i>et al</i> (2006)	Observational	10 000 patients for 3 years	10 European countries	Medication discontinuation and remission	Lowest discontinuation rate and highest remission rate with clozapine, then olanzapine, then risperidone	Pharmaceutical industry
Taylor <i>et al</i> (2006)	Observational	958 patients for up to 3 years	Scotland	Duration of treatment	Duration of treatment longest with clozapine, then (in rank order) olanzapine, risperidone, amisulpiride and quetiapine	Independent
Tiihonen <i>et al</i> (2006)	Observational	2230 first-episode patients up to 7 years	Finland	Discontinuation and hospitalisation rates	Lowest relapse with oral medication for clozapine, then (in rank order) olanzapine, thioridazine, perphenazine, risperidone and chlorpromazine	Government
Jones <i>et al</i> (2006)	RCT	227 for 56 weeks	England	Quality of life and symptoms	No difference between first- and second-generation antipsychotics	Government
Lieberman <i>et al</i> (2005)	RCT	1493 patients up to 18 months	USA	Medication discontinuation	Olanzapine most effective. No difference between other study medication	Government
McEvoy <i>et al</i> (2006)	RCT	400 first-episode patients for 1 year	USA	Duration of treatment	No difference between olanzapine, quetiapine and risperidone	Pharmaceutical industry
McCue <i>et al</i> (2006)	Pragmatic	Hospitalised patients for at least 3 weeks	USA	Hospital discharge and BPRS	Haloperidol, olanzapine and risperidone more effective than aripiprazole, quetiapine and ziprasidone	Independent

RCT, randomised controlled trial; BPRS, Brief Psychiatric Rating Scale.

whether discontinuation or rehospitalisation was considered, and this is echoed in the SOHO study (Haro *et al*, 2006). Another long-term study of over 500 patients in England (Hodgson *et al*, 2005) demonstrated the same rank order of effectiveness of oral atypicals using medication discontinuation as an outcome. In this study it was apparent that clozapine was being used for a treatment-resistant cohort. Taylor *et al* (2006) studied duration of treatment as a proxy in a Scottish population over 3 years and reported similar results to Tiihonen *et al* (2006) and Hodgson *et al* (2005).

McCue *et al* (2006) in a randomised open-label study of atypical antipsychotics and haloperidol in in-patients using the Brief Psychiatric Rating Scale (BPRS; Overall & Gorham, 1962) and time to discharge as outcome measures found similar effectiveness between haloperidol, olanzapine and risperidone and that these drugs were significantly better than aripiprazole and quetiapine. However, there was a

dissonance between time to discharge and the BPRS outcomes, which might suggest that rating instruments are not sensitive to important changes that influence management, at least in the short term. Although haloperidol was equal to risperidone and olanzapine it was associated with more extrapyramidal side-effects. Jones *et al* (2006) failed to detect any differences in effectiveness between first- and second-generation antipsychotics and reported no difference in extrapyramidal-type side-effects, in stark contrast to many other RCTs. A recent RCT of 400 first-episode patients (McEvoy *et al*, 2006) compared olanzapine, quetiapine and risperidone over 1 year and failed to detect a difference in discontinuation rates between these drugs although olanzapine had a significantly greater effect on positive symptoms. Discontinuation was associated with poor response ($P < 0.001$) and poor medication adherence ($P = 0.02$).

In general, RCTs are powered for one primary outcome which does not always

reflect primary clinical concern (McQuade *et al*, 2004). As observational studies are larger, there is more scope for legitimate subgroup analysis, such as treatment effect on those with comorbid disorder. The 3-year results of the SOHO study provide insights into social function and factors associated with relapse and remission. These are consonant with other independent studies and increase the face validity of this study. Although the SOHO study demonstrates relatively high switching rates for some medications, 65% of patients achieved remission, which resonates with the results of other long-term studies (Ciompi, 1980; Harding, 1988).

Observational studies and safety

Although often not acknowledged as such, post-marketing surveillance is essentially an observational study, albeit often poorly conducted (Vray *et al*, 2005). However, post-marketing surveillance often reports important safety information that was not

apparent from RCTs. The association between blood dyscrasias, clozapine and remoxipride are prime examples. In general, RCTs provide useful information on common adverse events, but identifying the relative risk of uncommon adverse events is realistically possible only in observational trials. In this regard, adverse event reporting in observational trials has been shown to enhance safety during the trial and facilitate the role of data monitoring committees and institutional review boards confronted with multiple reports of adverse events (Califf & Lee, 2001).

COMMON METHODOLOGICAL ISSUES

Analysis

Both RCTs and observational studies present difficulties in analysis. In RCTs high attrition rates have led to intention-to-treat analyses with a variety of statistical techniques evolving to accommodate these drop outs. These include last-observation-carried-forward (LOCF) analysis and mixed model repeated measures (MMRM); LOCF assumes that data are missing completely at random and that the patient's condition would remain constant; both assumptions are unlikely; MMRM is valid under less restrictive assumptions with use of missing data dependent on other measured factors (Mallinckrodt *et al*, 2003).

Randomised controlled trials have highlighted relatively high switching rates between therapies and potentially confounding baseline variation, with lower rates measured in observational studies. Baseline variation can be accommodated in analysis but, as with drop out from RCTs, it cannot be assumed that this variation is random and may reflect clinical practice. For example, in the study reported by Hodgson *et al*, (2005) and the SOHO study (Haro *et al*, 2006) young men with multiple illness episodes were more likely to receive clozapine.

Switching treatments within an observational study can be studied using marginal structural models (MSM), a new class of causal models that allow for improved adjustment of confounding in longitudinal data analysis in naturalistic settings by consistently estimating the parameters of the inverse-probability-of-treatment weighted estimators (Mortimer *et al*, 2005); MSM are an extension of propensity scoring to longitudinal data. Whereas propensity

scoring controls for selection bias by re-weighting observations to produce 'balance' between groups, MSM do the same but in a longitudinal fashion; MSM allow estimation of the causal effect of treatments in longitudinal naturalistic data when patients switch or stop treatment, even in the presence of missing (at random) data and time-varying confounding variables.

Patient concordance and sample size

In estimating treatment effects both RCTs and observational studies are challenged by patient concordance. Drug levels, which are highly variable for many psychotropics, are not routinely used, with pill counting being a common concordance measure in RCTs. However, poor adherence may underestimate treatment effects. Patient and clinician choice is important in determining outcome (Black, 1996) and controlling for these variables in RCTs limits the exploration of these factors. Zelen (1979) has advocated a methodology that has the advantage that, before providing consent, a patient will know whether an experimental treatment is to be used. Further development of patient and clinician preference trials has been described (Korn & Baumrind, 1991; Wennberg *et al*, 1993). McCue *et al* (2006) demonstrate that physician knowledge of a treatment might enhance optimum treatment dosing.

The nature of observational studies allows large sample sizes that add to the power of the study, facilitate subgroup analysis and provide data for robust sample size estimates for RCTs. Although in general appropriate sample sizes are important in RCTs, the superiority of those with large sample sizes over those with smaller samples has been challenged with regard to overestimating treatment effects (Contopoulos-Ioannidis *et al*, 2005).

Publication bias and sponsorship

Publication bias might also affect the two methodologies. Given the hierarchy of evidence, journals may be less willing to accept observational studies (Barton, 2000). Journals are less likely to publish negative studies and both methodologies are potentially biased by the study sponsor, with positive results often being associated with the vested interest of the sponsor (Als-Nielsen *et al*, 2003). However, a review of atypical antipsychotic trials and funding sources indicates that this is

not invariably so (Heres *et al*, 2006). Moreover, government-funded trials cannot be assumed to be unbiased (Coynne, 2006)

THE WAY FORWARD

The pre-eminence of RCTs and regulatory requirements has led to maintenance of the status quo in clinical drug trial development. Once a drug receives its marketing authorisation then further trial work is often aimed at developing markets rather than ascertaining whether the drug is effective. These concerns are just as relevant to psychotherapy and other non-pharmacological interventions. Making the trials as much like routine practice as possible may help to make RCTs more feasible and enhance external validity (so-called pragmatic trials; Hotopf, 2002). Although pragmatic trials may eschew some features of RCTs, such as double blinding, careful consideration may significantly reduce bias (Schulz *et al*, 1995). Patient recruitment is broad and may not be diagnostically driven (e.g. frequent attendees at a general practitioner surgery or people who self-harm). Outcomes, such as a reduction in suicide or episodes of violence, are clinically significant. Patient preference is an important variable in treatment choice which is negated in a traditional RCT, but patient preference trials have been reported (Ward *et al*, 2000) and may be particularly relevant when masking is not possible. The CATIE study (Lieberman *et al*, 2005) has many features of a pragmatic trial, such as narrow exclusion criteria and medication discontinuation as an outcome.

Randomised controlled trials and observational studies are not mutually exclusive, and there are examples from other areas of medicine of two designs running in parallel. For example, several studies quoted in Benson & Hartz (2000) in coronary artery disease illustrate the merits of enhancing an RCT by the addition of observational data from a concurrent registry of all non-randomised patients in the same centres. This approach improves the quality of observational research, since the same rigorous attention to detail in defining eligible patients, maintaining follow-up and recording outcomes is applied in both the randomised and the observational cohorts. The observational cohort may still suffer from selection bias, but there is a greater likelihood that its causes can be identified. The corollary also applies in that the

observational cohort inform on the typicality of the experimental group.

Rapid changes in methodologies without bridging links with older methodologies may preclude legitimate comparison and subsequent meta-analysis. However, advances in the understanding of the biological and psychological mechanisms of mental illness will also dictate the evolution of relevant end-points. This is typified by the increasing interest in cognitive outcomes (Stroup *et al*, 2003) for which NICE recommends audits and provides standardised templates. This is another potential for supplementing treatment information and should facilitate the collection of data pools that inform treatment practice. The introduction of new treatment presents the possibility of mirror image studies (Hodgson *et al*, 2002) that allow some measure of utility, although regression towards the mean precludes overinterpretation of the results.

CONCLUSIONS

The RCT has served medicine well but evaluation of treatment needs reviewing for the 21st century. Outcomes need to be more clinically relevant and comparable with those from other trial methodologies. Biases in recruitment need to be addressed and post-marketing surveillance needs a more robust approach, as does monitoring of fidelity to treatment or service delivery models. In part this could be achieved with naturalistic studies, audits and mirror image studies. Without such additional information, treatments cannot be tailored effectively to the patient. Dogma should not be allowed to drive the experimental paradigm agenda as no current research design provides comprehensive clinical information.

ACKNOWLEDGEMENTS

R. Hunter has received funding from NHS Quality Improvement Scotland and the Chief Scientist Office, Edinburgh.

REFERENCES

- Ackner, B. & Oldham, J. A. (1960)** Insulin treatment of schizophrenia: a controlled study. *Lancet*, *i*, 711.
- Als-Nielsen, B., Chen, W., Gluud, C., et al (2003)** Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*, **290**, 921–928.
- Barton, S. (2000)** Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *BMJ*, **321**, 255–256.
- Benson, K. & Hartz, A. J. (2000)** A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine*, **342**, 1878–1886.
- Black, N. (1996)** Why we need observational studies to evaluate the effectiveness of health care. *BMJ*, **312**, 1215–1218.
- Boardman, A. P., Hodgson, R. E., Lewis, M., et al (1999)** The North Staffordshire Community Beds Study: longitudinal evaluation of psychiatric in-patient units attached to community mental health centres. I. Methods, outcome and patient satisfaction. *British Journal of Psychiatry*, **175**, 70–78.
- Bonell, C., Oakley, A. & Hargreaves, J. (2006)** Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ*, **333**, 346–349.
- Burns, T. (2006)** NICE guidance in schizophrenia: how generalisable are drug trials? *Psychiatric Bulletin*, **30**, 210–212.
- Bushe, C. J. & Shaw, M. (2007)** Prevalence of hyperprolactinaemia in a naturalistic cohort of schizophrenia and bipolar outpatients during treatment with typical and atypical antipsychotics. *Journal of Psychopharmacology*, in press. doi: 10.1177/0269881107078281.
- Califf, R. M. & Lee, K. L. (2001)** Data and safety monitoring committees: philosophy and practice. *American Heart Journal*, **141**, 154–155.
- Ciampi, L. (1980)** The natural history of schizophrenia in the long term. *British Journal of Psychiatry*, **136**, 413–420.
- Concato, J., Shah, N., & Horwitz, R. I. (2000)** Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, **342**, 1887–1892.
- Contopoulos-Ioannidis, D. G., Gilbert, S. M., Trikalinos, T. A., et al (2005)** Comparison of large versus smaller randomized trials for mental health-related interventions. *American Journal of Psychiatry*, **162**, 578–584.
- Cook, T. D. & Campbell, D. T. (1979)** *Quasi-Experimentation: Design and Analysis for Field Settings*. Houghton Mifflin.
- Coyne, J. C. (2006)** Cochrane reviews v industry supported meta-analyses. We should read all reviews with caution. *BMJ*, **333**, 916.
- Csernansky, J. G., Mahmoud, R. & Brenner, R. (2002)** A comparison of risperidone and haloperidol for the prevention of relapse in patients with schizophrenia. *New England Journal of Medicine*, **346**, 16–22.
- Davis, J. M., Chen, N., & Glick, I. D. (2003)** A meta analysis of the efficacy of second generation antipsychotics. *Archives of General Psychiatry*, **60**, 553–564.
- Dean, C., Phillips, J., Gadd, E. M., et al (1993)** A comprehensive community based service for people with acute severe episodes of illness. *BMJ*, **307**, 473–476.
- El-Sayeh, H. G., Morganti, C. & Adams, C. E. (2006)** Aripiprazole for schizophrenia. Systematic review. *British Journal of Psychiatry*, **189**, 102–108.
- Harding, C. M. (1988)** Course type in schizophrenia: an analysis of European and American studies. *Schizophrenia Bulletin*, **14**, 633–643.
- Haro, J. M., Kamath, S. A., Ochoa, S., et al (2003)** The Clinical Global Impression – Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatrica Scandinavica*, **107**, 16–23.
- Haro, J. M., Novick, D. & Suarez, D. (2006)** Remission and relapse in the outpatient care of schizophrenia. *Journal of Clinical Psychopharmacology*, **26**, 571–578.
- Harrison-Read, P., Lucas, B., Tyrer, P., et al (2002)** Heavy users of acute psychiatric beds: randomized controlled trial of enhanced community management in an outer London borough. *Psychological Medicine*, **32**, 403–416.
- Heres, S., Davis, J., Maino, K., et al (2006)** Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *American Journal of Psychiatry*, **163**, 185–194.
- Hoblyn, J., Noda, A., Yesavaga, J. A., et al (2006)** Factors in choosing atypical antipsychotics: toward understanding the bases of physicians' prescribing decisions. *Journal of Psychiatric Research*, **40**, 160–166.
- Hodgson, R. E. (2005)** Long term outcomes and atypical antipsychotic discontinuation rates. Do different methodologies give similar results? *European Neuropsychopharmacology*, **15** (suppl. 3), 467.
- Hodgson, R. E. & Belgamwar, R. (2006)** Off-label prescribing by psychiatrists. *Psychiatric Bulletin*, **30**, 55–57.
- Hodgson, R. E., Lewis, M. & Boardman, A. P. (2001)** The prediction of readmission to acute psychiatric wards. *Social Psychiatry and Psychiatric Epidemiology*, **36**, 304–309.
- Hodgson, R. E., Carr, D. & Wealleans, L. (2002)** Brunswick House: a weekend crisis house in North Staffordshire. *Psychiatric Bulletin*, **26**, 453–455.
- Hodgson, R. E., Belgamwar, R., Al-tawarah, Y., et al (2005)** The use of atypical antipsychotics in the treatment of schizophrenia in North Staffordshire. *Human Psychopharmacology: Clinical and Experimental*, **20**, 141–147.
- Hotopf, M. (2002)** The pragmatic randomised controlled trial. *Advances in Psychiatric Treatment*, **8**, 326–333.
- Ioannidis, J. P. & Lau, J. (2001)** Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA*, **285**, 437–443.
- Jones, P. B., Barnes, T. R. E., Davies, L., et al (2006)** Randomized controlled trial of the effect on quality of life of second- vs first-generation antipsychotic drugs in schizophrenia: cost utility of the latest antipsychotic drugs in schizophrenia study (CUtLASS I). *Archives of General Psychiatry*, **63**, 1079–1087.
- Kane, J. M., Honigfeld, G. & Singer, J. (1988)** Clozapine in the treatment-resistant schizophrenic: a double-blind comparison with chlorpromazine. *Archives of General Psychiatry*, **45**, 789–796.
- Kasper, S., Rosillon, D. & Duchesne, I. (2001)** Risperidone Olanzapine Drug Outcomes studies in Schizophrenia (RODOS): efficacy and tolerability results of an international naturalistic study. *International Clinical Psychopharmacology*, **16**, 179–187.
- Kinon, B. K., Liu-Seifert, H., Adams, D. H., et al (2006)** Differential rates of treatment discontinuation as a measure of treatment effectiveness for olanzapine and comparator antipsychotics for schizophrenia. *Journal of Clinical Psychopharmacology*, **26**, 632–637.
- Korn, E. L. & Baumrind, S. (1991)** Randomised clinical trials with clinician-preferred treatment. *Lancet*, **337**, 149–152.
- Lenzer, J. (2004)** Pfizer pleads guilty, but drug sales continue to soar. *BMJ*, **328**, 1217.

Lester, H. & Wilson, S. (1999) Practical problems in recruiting patients with schizophrenia into randomised controlled trials. *BMJ*, **318**, 1075.

Leucht, S., Barnes, T. R. E., Kissling, W., et al (2003) Relapse prevention in schizophrenia with new-generation antipsychotics: a systematic review and exploratory meta-analysis of randomized, controlled trials. *American Journal of Psychiatry*, **160**, 1209–1222.

Leucht, S., Kane, J. M., Kissling, W., et al (2005) What does the PANSS mean? *Schizophrenia Research*, **79**, 231–238.

Lieberman, J. A., Phillips, M., Gu, H., et al (2003) Atypical and conventional antipsychotic drugs in treatment-naïve first-episode schizophrenia: a 52-week randomized trial of clozapine vs chlorpromazine. *Neuropsychopharmacology*, **28**, 995–1003.

Lieberman, J. A., Stroup, T. S., McEvoy, J. P., et al (2005) Clinical antipsychotic trials of intervention effectiveness (CATIE) Investigators: Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*, **353**, 1209–1223.

Linden, A., Adams, J. L. & Roberts, N. (2006) Strengthening the case for disease management effectiveness: un-hiding the hidden bias. *Journal of Evaluation in Clinical Practice*, **12**, 140–147.

Mallinckrodt, C. H., Sanger, T. M., Dube, S., et al (2003) Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, **15**, 754–760.

McCue, R. E., Waheed, R., Urcuyo, L., et al (2006) Comparative effectiveness of second-generation antipsychotics and haloperidol in acute schizophrenia. *British Journal of Psychiatry*, **189**, 433–440.

McEvoy, J. P., Perkins, D. O., Gu, H., et al (2006) Olanzapine, quetiapine, and risperidone in the treatment of first-episode psychosis: effectiveness and factors influencing adherence to treatment. *European Neuropsychopharmacology*, **16** (suppl. 4), S425–426.

McQuade, R. D., Stock, E., Marcus, R., et al (2004) A comparison of weight change during treatment with olanzapine or aripiprazole: results from a randomized, double-blind study. *Journal of Clinical Psychiatry*, **65** (suppl. 18), 47–56.

Meehan, K., David, S., Tohen, M., et al (2001) A double blind randomised comparison of the efficacy and safety of intramuscular injections of olanzapine, lorazepam or placebo in treating acutely agitated patients diagnosed with bipolar mania. *Journal of Clinical Psychopharmacology*, **21**, 389–397.

RICHARD HODGSON, MB ChB, MSc, MD, MRCPsych, Lyme Brook Centre, Stoke on Trent; CHRIS BUSHE, MB BS, Eli Lilly, Lilly House, Basingstoke; ROBERT HUNTER, BSc, MD, FRCPsych, University Department of Psychological Medicine, Gartnavel Royal Hospital, Glasgow, UK

Correspondence: Dr Richard Hodgson, Lyme Brook Centre, Bradwell Hospital, Talke Road, Stoke-on-Trent, Staffordshire ST5 7TL, UK. Email: richarde.hodgson@northstuffs.nhs.uk

Moncrieff, J. (1997) Lithium: evidence reconsidered. *British Journal of Psychiatry*, **171**, 113–119.

Mortimer, K. M., Neugebauer, R., van der Laan, M., et al (2005) An application of model-fitting procedures for marginal structural models. *American Journal of Epidemiology*, **15**, 382–388.

National Institute for Clinical Excellence (2002) *Guidance on the Use of Newer (Atypical) Antipsychotic Drugs for the Treatment of Schizophrenia*. NICE.

Overall, J. E. & Gorham, D. R. (1962) The Brief Psychiatric Rating Scale. *Psychological Report*, **10**, 799–812.

Papanikolaou, P. N., Churchill, R., Wahlbeck, K., et al (2004) Safety reporting in randomized trials of mental health interventions. *American Journal of Psychiatry*, **161**, 1692–1697.

Phelan, M., Stradins, L. & Morrison, S. (2001) Physical health of people with severe mental illness. *BMJ*, **322**, 443–444.

Roland, M. & Torgerson, D. J. (1998) What are pragmatic trials? *BMJ*, **316**, 285.

Schulz, K. F., Chalmers, I., Hayes, R. J., et al (1995) Empirical evidence of bias. *JAMA*, **273**, 408–412.

Stroup, T. S., McEvoy, J. P. & Swartz, M. S. (2003) The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Project: schizophrenia trial design and protocol development. *Schizophrenia Bulletin*, **29**, 15–31.

Taylor, A. M., Shajahan, P. & Carleton, R. (2006) Comparing the use and outcomes of antipsychotics in the real world. *European Neuropsychopharmacology*, **16** (suppl. 4), S414–415.

Thornicroft, G. & Strathdee, G. (1994) How many psychiatric beds. *BMJ*, **309**, 970–971.

Thornley, B. & Adams, C. (1998) Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ*, **317**, 1181–1184.

Tiihonen, J., Wahlbeck, K., Lönnqvist, J., et al (2006) Effectiveness of antipsychotic treatments in a nationwide cohort of patients in community care after first hospitalisation due to schizophrenia and schizoaffective disorder: observational follow-up study. *BMJ*, **333**, 224.

van Os, J., Drukker, M. A., Campo, J., et al (2006) Validation of remission criteria for schizophrenia. *American Journal of Psychiatry*, **163**, 2000–2002.

von Knorring, L. & Lindstrom, E. (1995) Principal components and further possibilities with the PANSS. *Acta Psychiatrica Scandinavica Supplementum*, **388**, 5–10.

Vray, M., Hamelin, B. & Jaillon, P. (2005) The respective roles of controlled clinical trials and cohort monitoring studies in the pre- and postmarketing assessment of drugs. *Therapie*, **60**, 339–349.

Ward, E., King, M., Lloyd, M., et al (2000) Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care for patients with depression I: clinical effectiveness. *BMJ*, **321**, 1393–1399.

Wennberg, J. E., Barry, M. J., Fowler, F. J., et al (1993) Outcomes research, PORTS, and health care reform. Doing more good than harm: the evaluation of health care interventions. *Annals of the New York Academy of Science*, **703**, 56–62.

Williams, D. D. R. & Garner, J. (2002) The case against 'the evidence': a different perspective on evidence-based medicine. *British Journal of Psychiatry*, **180**, 8–12.

Wright, P., Birkett, M., David, S., et al (2001) Double blind, placebo-controlled comparison of intramuscular olanzapine and intramuscular haloperidol in the treatment of acute agitation in schizophrenia. *American Journal of Psychiatry*, **158**, 1149–1151.

Yastrubetskaya, O., Chiu, E. O. & Connell, S. (1997) Is good clinical research practice for clinical trials good for clinical practice? *International Journal of Geriatric Psychiatry*, **12**, 227–231.

Zelen, M. (1979) A new design for randomized clinical trials. *New England Journal of Medicine*, **300**, 1242–1245.