No medication without representation? Generalizing from antidepressant clinical efficacy trials to clinical practice

A commentary on 'The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice' by van der Lem *et al.* (2010)

I. M. Anderson*

Neuroscience and Psychiatry Unit, School of Community Based Medicine, Faculty of Medical and Human Sciences, University of Manchester and Manchester Academic Health Science Centre, Manchester, UK

Introduction

It used to be rather simple. A patient presents with depression – prescribe an antidepressant based on evidence from clinical trials against placebo going back to the 1960s. Now, however, there is increasing questioning of the benefits of antidepressants due to a range of factors ranging at one end from distrust of 'big Pharma' and concerns about medicalizing distress (Moncrieff, 2008) through to questioning how well antidepressants actually work in the treatment of depression (e.g. Kirsch *et al*. 2008; Pigott *et al*. 2010). The last is a deceptively simple question that is so hard to answer it raises the suspicion that there is something not quite right with its formulation. Depression is defined at a syndromal level but it is difficult to get away from the fact that patients who meet the syndromal diagnosis are heterogeneous. Randomized controlled clinical trials to test antidepressant efficacy (antidepressant efficacy trials; AETs) in contrast aim for a homogeneous and low-risk patient group, usually in order to obtain results that can be used for licensing purposes. One of the problems about asking how well antidepressants work is in defining the population we mean. Is it the population/s studied in AETs or the one/s seen in the clinic or community – do they differ, by how much, and does it matter in terms of outcome? If there are major differences, and it does matter, then the almost exclusive reliance on

AETs to inform clinical treatment guidelines could pose a major problem.

Does AET eligibility affect outcome?

In their paper van der Lem *et al*. (2010) set out to address this issue by studying a large cohort of depressed patients treated in secondary care in the Netherlands. They compare the outcome in those who would have been eligible for a 'typical' AET with those who would have been excluded from such a trial. In agreement with other studies (e.g. Zimmerman *et al*. 2002; Wisniewski *et al*. 2009) they found that only a minority of patients (17–25%, depending on the stringency of criteria) would have been eligible for an AET. The two most important factors in their cohort were severity of depression and co-morbid (mainly anxiety) disorders. The results in brief were that AET eligibility did not influence outcome 5 months later and factors involved in AET inclusion or exclusion (and the type of treatment) accounted for less than 5% of the variance in outcome. The authors conclude that the influence of eligibility for AETs on treatment outcome in clinical practice seems to be small.

This result appears at first sight to be at odds with those from the other main study to examine this, the large naturalistic Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study set in the USA (Wisniewski *et al*. 2009). In this study the AET-eligible, compared with -ineligible, group, had considerably higher rates of response (51.6% *v.* 39.1%) and remission (34.4% *v.* 24.7%) when they received open-label treatment with citalopram. There are, however, considerable differences between the two studies that

* Address for correspondence : Dr I. M. Anderson, Neuroscience and Psychiatry Unit, School of Community Based Medicine, Faculty of Medical and Human Sciences, University of Manchester and Manchester Academic Health Science Centre, Stopford Building, Oxford Road, Manchester M13 9PT, UK.

(Email : ian.anderson@manchester.ac.uk)

are likely to explain the apparent discrepancy. I shall concentrate on the two most important. First, patients in the STAR*D study were all treated with a single antidepressant agent over an average of 8 weeks in a standardized manner whereas those in the van der Lem *et al.* study had a variety of treatments (adequacy unknown) over 5 months and only just over half (54%) of those followed up received antidepressants. Second, there was a high drop-out rate in the van der Lem *et al.* study, with only 46% having a follow-up assessment compared with over 92% in the STAR*D study. Although there did not appear to be systematic baseline differences between those followed up and those lost to follow-up the truth is, with such a large percentage lost, we just cannot be sure what happened to them and how representative the results are. Overall, these limitations make the van der Lem *et al.* study difficult to interpret with regard to telling us about the influence of AET eligibility on outcome, and certainly on response to any specific treatment.

Can clinical trials ever be generalizable?

One of the striking features from the van der Lem *et al.* study is an even worse outcome than found in the STAR*D study, with only 28% responding and 21% remitting in the former study compared with 47% responding and 28% remitting in the latter (Trivedi *et al.* 2006). This probably reflects another, almost certainly important, difference between the studies – their patient populations. The Netherlands has a primary-care gatekeeper system so that referral to secondary care will be 'filtered' by factors such as treatment non-response, risk or treatment availability, making the patients not representative of depressed patients as a whole, or even those presenting to primary care. The USA does not have primary-care gatekeepers but the study required explicit consent from patients to participate in a clinical trial and there were exclusion criteria based on co-morbidity and previous treatment response (although not as strict as for an AET) so that only 65% of those agreeing to be screened entered into the study (Trivedi *et al.* 2006). Therefore the STAR*D patients were also only a subgroup of those seen in clinical practice, but a different subgroup to the Netherlands sample.

Where does this leave us regarding the question as to whether we can generalize from AETs to clinical practice? We can be certain that patients in AETs are not representative overall of depressed patients seen in clinical practice (the problem of what is a representative sample is of course another matter as it will vary by setting and healthcare system). The STAR*D study is the best evidence that we have so far that AET eligibility does influence treatment response but this

should be no surprise given what we already know about predictors of outcome (O'Leary *et al.* 2000; Van Henricus *et al.* 2008; Kim *et al.* 2011). What we cannot tell from studies such as this is the degree to which we can use the results from AETs to guide practice. The efficacy rationale for antidepressants is based on separation from placebo, not overall outcome. From the limited evidence available the factors that influence response to antidepressants may also influence response to placebo in AETs (Stewart *et al.* 1989; Angst *et al.* 1993). The truth is that we just do not know whether the pharmacological effect (i.e. effect size against placebo) that we see in AETs reflects what occurs in other patient populations. The disappointingly poor response in these more general populations does, however, emphasize that antidepressants on their own have limited benefits and that other approaches are required.

Matching patient to treatment

We will never be able to ethically randomize a fully representative patient sample to placebo-controlled, or probably even comparator-controlled, trials (assuming we know what we mean by representative). Even if we could, patients will never be treated in clinical practice in the same way they are in a randomized clinical trial. On the other hand, the more naturalistic and representative a study is, the more outcome will affected by a myriad of factors, making the contribution of an individual treatment hard to detect (in the absence of large sustained effects and 'hard' objective outcomes such as death or hospitalization). The global question as to whether we can extrapolate AETs to clinical practice therefore becomes somewhat meta-physical in that we can never answer it. AETs can provide evidence for a pharmacological antidepressant effect of a drug in groups of patients with syndromally defined depression but over-interpretation and over-generalization of the precision of the size of effect to an individual depressed patient seen in a clinical setting makes little sense. AETs do provide useful, but limited, information about when it may be helpful to treat with an antidepressant, what to use and the likely duration of benefit (Anderson *et al.* 2008). However, applying group effects to individual patients is always going to be inherently uncertain, and the less eligible a patient is for an AET the less certain we are able to be. This uncertainty and wide individual variation in response to specific treatments have led to current interest in whether or not it may be possible to personalize treatment for depression based on developing predictors for response such as sociodemographic and clinical characteristics, and biological markers (e.g. neuroimaging or genetic variation)

(Simon & Perlis, 2010). Currently this variability is unpredictable, and if largely stochastic like the weather could remain so, but given the poor outcomes for depression we owe it to our patients to try and find useful predictors. At present, however, good clinical treatment of depressed patients has to remain a skilful art which uses best evidence to inform individualized treatment trials guided by careful evaluation of outcome, not forgetting that it needs to be in the context of full clinical assessment, patient education and negotiation and integration of psychosocial and drug treatment approaches.

Declaration of Interest

I.M.A. has received grant support and honoraria for speaking from pharmaceutical companies marketing antidepressants and chaired the Development Guideline Group for the National Institute for Health and Clinical Excellence Depression Guideline Update.

References

Anderson IM, Ferrier IN, Baldwin RC, Cowen PJ, Howard L, Lewis G, Matthews K, McAllister-Williams RH, Peveler RC, Scott J, Tylee A (2008). Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *Journal of Psychopharmacology* **22**, 343–396.

Angst J, Scheidegger P, Stabl M (1993). Efficacy of moclobemide in different patient groups. Results of new subscales of the Hamilton Depression Rating Scale. *Clinical Neuropharmacology* **16** (Suppl. 2), S55–S62.

Kim JM, Kim SW, Stewart R, Kim SY, Yoon JS, Jung SW, Lee MS, Yim HW, Jun TY (2011). Predictors of 12-week remission in a nationwide cohort of people with depressive disorders: the CRESCEND study. *Human Psychopharmacology*. Published online: 23 February 2011. doi:10.1002/hup.1168.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine* **5**, e45.

Moncrieff J (2008). The creation of the concept of an antidepressant: an historical analysis. *Social Science and Medicine* **66**, 2346–2355.

O'Leary D, Costello F, Gormley N, Webb M (2000). Remission onset and relapse in depression. An 18-month prospective study of course for 100 first admission patients. *Journal of Affective Disorders* **57**, 159–171.

Pigott HE, Leventhal AM, Alter GS, Boren JJ (2010). Efficacy and effectiveness of antidepressants: current status of research. *Psychotherapy and Psychosomatics* **79**, 267–279.

Simon GE, Perlis RH (2010). Personalized medicine for depression: can we match patients with treatments? *American Journal of Psychiatry* **167**, 1445–1455.

Stewart JW, McGrath PJ, Quitkin FM, Harrison W, Markowitz J, Wager S, Leibowitz MR (1989). Relevance of DMS-III depressive subtype and chronicity of antidepressant efficacy in atypical depression. Differential response to phenelzine, imipramine, and placebo. *Archives of General Psychiatry* **46**, 1080–1087.

Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, Norquist G, Howland RH, Lebowitz B, McGrath PJ, Shores-Wilson K, Biggs MM, Balasubramani GK, Fava M (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *American Journal of Psychiatry* **163**, 28–40.

van der Lem R, van der Wee NJA, van Veen T, Zitman FG (2010). The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. *Psychological Medicine*. Published online: 16 November 2010. doi:10.1017/S0033291710002175.

Van Henricus L, Schoevers RA, Dekker J (2008). Predicting the outcome of antidepressants and psychotherapy for depression: a qualitative, systematic review. *Harvard Review of Psychiatry* **16**, 225–234.

Wisniewski SR, Rush AJ, Nierenberg AA, Gaynes BN, Warden D, Luther JF, McGrath PJ, Lavori PW, Thase ME, Fava M, Trivedi MH (2009). Can phase III trial results of antidepressant medications be generalized to clinical practice? A STAR*D report. *American Journal of Psychiatry* **166**, 599–607.

Zimmerman M, Mattia JI, Posternak MA (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry* **159**, 469–473.