9

What Antidepressant Research Really Shows

Overwhelmingly, the most common response to our serotonin paper was that it doesn't matter because 'antidepressants work'. In the online magazine *The Conversation*, two UK-based psychiatrists wrote an article entitled 'Depression: low serotonin may not be the cause – but antidepressants still work' and an Australian psychiatrist published one called 'The chemical imbalance theory of depression is dead, but that doesn't mean antidepressants don't work'. Both pieces made the same argument: it's not important that we don't understand *how* antidepressants work, only that they do.[1]

So, before I cover the reactions to our paper and what they reveal about how public ideas about depression and its treatment are shaped and maintained, it is time to take a detailed look at the evidence for whether antidepressants actually help people. This involves presenting some dry facts about drug trials and how they are conducted and interpreted, but it is essential that people understand exactly what the much-repeated claim that 'antidepressants work' is based on. The issues are not difficult to grasp, so bear with me and hopefully the effort will be worthwhile. When you have read the next few chapters, you will be more informed than most doctors.

I hope it is already clear that how antidepressants exert the effects they do matters a lot, but whether they work is also doubtful. When I was training as a psychiatrist, I saw many people with depression who were starting out on antidepressants. Some got better, some didn't. Other

psychiatrists and many patients thought the antidepressants were helpful, but I just couldn't see this. When people improved, usually the things that had made them depressed in the first place had resolved or improved – they had got a new job or recovered from a break-up, for example. I realised people can interpret the same situation in different ways. Many people are inclined to believe treatment is beneficial and they attribute improvements to treatment when there might be other explanations.

This is why we do drug trials. It is why we need to compare drugs like antidepressants to a placebo – that is, a 'dummy' tablet. We know that our moods go up and down naturally; we know moods reflect our circumstances; and we know that being seen by a professional – being listened to and sympathised with – all enhance people's chances of improvement.[2] Taking a pill we think *might* be effective can also make us feel better. Therefore, we need to know whether taking an antidepressant is more effective than not taking one, and, if it is, whether the effects are due to a specific pharmacological effect of the drug or whether they are down to people's beliefs about the benefits of the drug – what we call the placebo effect.

Most drug trials are what are called 'randomised-controlled trials'. This means people who enrol are assigned at random to take the active drug or the placebo. The randomness of the process ensures that there are no systematic differences between people allocated to the different sorts of tablets (the drug and the placebo tablets) that could influence the outcome of the trial.

When I was a junior doctor, I worked on a drug trial testing whether a particular drug (naltrexone) was helpful for people with alcohol problems. I was employed by an NHS hospital in London, but the trial was organised by a clinical research organisation (a company that conducts drug trials), which had been hired by the drug's maker – the pharmaceutical company DuPont. My job consisted of identifying patients who might be eligible for the trial, signing them up if they were willing, and then conducting medical examinations and taking blood for tests. My overriding memory is the

inconvenience of having to lug round a large machine to spin the blood samples before I sent them off to the laboratory.

The experience gave me an invaluable insight into how trials work on the ground. I learned, for example, that most people who sign up for a trial want to get the real drug, not the placebo. Many of the people I recruited were desperate and would try anything that might help them stay sober. They were also intensely curious about the identity of their tablets and tried to guess whether they were taking the active drug or the placebo. We'll come to why this is important shortly.

Results of antidepressant trials

Many hundreds of trials have been conducted that involve comparing an antidepressant with a dummy tablet or placebo. Most have been run by pharmaceutical companies and typically last around six to eight weeks. The main thing that is measured is the level of symptoms people experience, which is assessed using questionnaires that ask people questions about their feelings and other symptoms of depression, anxiety or whatever condition is being treated.

So, what do these trials show? Because there are so many of them, the results of individual trials are commonly combined using the technique of meta-analysis. In depression, meta-analyses indicate that, overall, people assigned to take an antidepressant show slightly more improvement during the course of the trial than people assigned to take a placebo. The difference is 'statistically significant', meaning it is not just a chance finding,[3] but it is small.

The most commonly used depression questionnaire was designed by a psychiatrist called Max Hamilton back in the 1960s and is called the Hamilton Depression Rating Scale.[4] The usual version has seventeen questions and a maximum score of 52 points. Meta-analyses show that the

average difference on this scale between people randomised to take an antidepressant and those randomised to take a placebo is 2 points or less.

A meta-analysis published in 2022, for example, which was based on data from antidepressant trials submitted to the US drug regulator, the FDA, found an average difference of 1.8 points.[5] A large and influential meta-analysis published in 2018, by a team based at the University of Oxford, revealed an overall difference of 2 points (although this was not how the results were presented).[6]

The relevance of antidepressant-placebo differences

So, the next question is, what does this mean? Is a two-point difference between an antidepressant and a placebo a relevant and helpful difference, and does it reflect a pharmacological effect, or might it be due to something else? On the face of it, a 2-point difference on a 52-point scale doesn't sound like much, but what does the evidence suggest?

Measuring depression is not a simple task, of course. It is not like taking someone's blood pressure or measuring the amount of sugar in the blood. Depression-rating scales consist of questions about a number of manifestations or symptoms of depression. The Hamilton Depression Rating Scale, like other depression scales, is an arbitrary collection of some of these. It doesn't include a question on loss of pleasure in life, which is often considered a cardinal indication of depression, and it contains several questions about physical symptoms, such as gastrointestinal symptoms and menstrual disturbances, which might have nothing to do with a person's mood.

The way the items are rated is also questionable. Take the depressed mood item presented here, for example:

DEPRESSED MOOD (sadness, hopeless, helpless, worthless)

- **Absent**

1 These feeling states indicated only on questioning

2 These feeling states spontaneously reported verbally

3 Communicates feeling states non-verbally, i.e. through facial expression, posture, voice and tendency to weep

4 Patient reports virtually only these feeling states in his/her spontaneous verbal and non-verbal communication

You are considered to be more depressed if you spontaneously tell the person doing the assessment you are depressed (score of 2) than if you only admit to it after being questioned (score of 1). People taking both antidepressants and placebos in clinical trials usually score between 1 and 2 at the end of their treatment, so this is a critical distinction.[7] But people have different styles of communication, and no one has established that telling someone about your mood directly indicates you are more severely depressed than if you are more reserved. We might expect that sometimes, at least, it would be the opposite.

It is also not necessarily true that people who show their feelings non-verbally, by crying for example (and would score 3), are more affected than people who don't (and would score 1 or 2).

Some people have suggested the Hamilton Depression Rating Scale isn't a very good scale, which is true, and that other scales might reveal larger antidepressant effects – they don't. The other most commonly used scale is the Montgomery–Åsberg Depression Rating Scale (known as the MADRS), named after the two psychiatrists who developed it.[8] Trials using this scale find modest differences between antidepressants and placebo that are roughly equivalent to the 2-point difference on the Hamilton Depression Scale.[9]

The problem is, there is no objective way of identifying or measuring depression; therefore, we don't know whether these scales are really measuring what they say they are measuring. When it comes down to it, we

can’t be confident that someone who scores 20 points on a depression scale is more depressed than someone who scores 10 points, and it certainly makes no sense to say they are precisely twice as depressed.

Given this, you could be forgiven for thinking that antidepressant trials are complete nonsense – and I would say this is not far from the truth. But putting aside more fundamental concerns about the whole enterprise of measuring depression, what evidence do we have about what a 2-point difference in Hamilton Depression Scale scores might mean? Most efforts to establish this suggest it doesn’t represent a helpful or even noticeable difference.

One way you can evaluate the meaning of depression scores is to compare them to scores on a widely used rating scale called the Clinical Global Impressions Scale, which assesses how people are doing overall.[10] This scale requires a researcher to classify people into one of seven categories: ‘very much improved’, ‘much improved’, ‘mildly improved’, ‘no change’, ‘minimally worse’, ‘much worse’ or ‘very much worse’.

A group of researchers from Germany, led by the respected and prolific Stefan Leucht, analysed data from a trial in which people were rated using the Hamilton Depression Scale and the Clinical Global Impressions Scale at the same time. Their analysis revealed that a change of 3 points or less on the Hamilton Depression Scale equated to the ‘no change’ category on the Clinical Global Impressions Scale. A change of 7 or 8 points equated to a rating of ‘mildly improved’.[11] In other words, a difference of 2 points does not even register as a difference on the Clinical Global Impressions Scale and a difference of 7–8 points would be needed to indicate that antidepressants had even a ‘mild’ advantage over a placebo.

Various other methods have been proposed to establish the size of a meaningful difference in Hamilton Depression Scale scores. None of the methods are perfect, by any means, because of the difficulty of measuring something like a mood, but they also indicate that a difference of 2 points would not qualify.[12] So, antidepressants trials show that antidepressants are not meaningfully different from a placebo.

THE AMPLIFIED PLACEBO EFFECT

There is another major problem with interpreting antidepressant trials, however, which suggests the small difference between antidepressants and placebo may not even be a pharmacological effect at all. It may be an 'amplified' placebo effect.

When I was a junior psychiatrist wondering why my colleagues thought antidepressants were so effective, I asked Geoff, one of the more senior doctors on my team, what he thought. Geoff was a supportive mentor. He always had good advice on how to manage situations in which people were suicidal, acutely psychotic or heavily intoxicated – the daily challenges of my work on the night shift in a busy accident and emergency department in London. Geoff said he thought antidepressants were 'active' placebos, and he referred me to a paper published in 1982 in the *British Journal of Psychiatry*. Reading the paper was a lightbulb moment for me. Everything fell into place. Now I understood why antidepressants might look a little bit better than a placebo in a clinical trial but had no convincing effect in the real world.

The paper Geoff signposted me to was called 'Side effects and placebo amplification' and was written by a psychiatrist called Richard Thomson.[13] Thomson pointed out how placebo-controlled trials do not necessarily eliminate the placebo effect. This is because antidepressants are active drugs and, therefore, the experience of taking them is different from the experience of taking an inert substance, such as chalk or lactose – the usual constituents of placebo tablets. Antidepressants have recognisable side-effects, such as nausea and drowsiness, and they are psychoactive drugs that change people's feelings and sensations in more- or less-subtle ways. Based on these effects, it is likely that at least some people taking part in antidepressant trials will be able to figure out whether they are taking the antidepressant or the placebo. The staff working on trials may also be able to guess who is taking the antidepressant and who is on the placebo by the profile of side-effects that participants report.

The point of running a randomised placebo-controlled trial is that none of the people involved, neither the participants nor the researchers, know who is getting the real drug and who is not. They are meant to be ‘blind’ to the nature of the tablets. This is what is known as the ‘double-blind’ design. But if people can guess what they are taking, or if the researchers can guess who is taking what, then the study is not double-blind.

In this situation, the results may be influenced by people’s common belief that the drug will be effective. Those who guess they are taking the real thing will be more optimistic and hopeful and this may improve their mood in itself. Conversely, those who suspect they are on the placebo may feel disappointed and dejected, which will make their mood worse. Researchers often have the same expectations, and they might rate people they suspect to be taking the real drug as doing better than those they think are taking the placebo.

Therefore, when a trial is not properly double-blind, and participants and researchers can guess who received the drug and who received the placebo, all or part of the effects of the drug may be due to an ‘amplified’ placebo effect. This is a combination of the ordinary placebo effect, which is the consequence of taking some sort of tablet, amplified by the extra boost people derive from the side-effects and other effects that suggest to people they are getting the real drug. In this context, the drug acts like a placebo with special powers.

This problem was recognised back in the early days of psychiatric drug research in the 1960s. To try to address it, several trials were set up that compared an antidepressant with an ‘active’ placebo, which is a drug that produces some of the same side-effects as the antidepressant being tested, without being thought to be an antidepressant. Thomson described seven such studies conducted in the 1960s and 1970s, which used the drug atropine as an active placebo in trials of various tricyclic antidepressants that were being used at the time. Atropine mimics some, though not all, of the side-effects of tricyclic antidepressants, particularly a dry mouth.

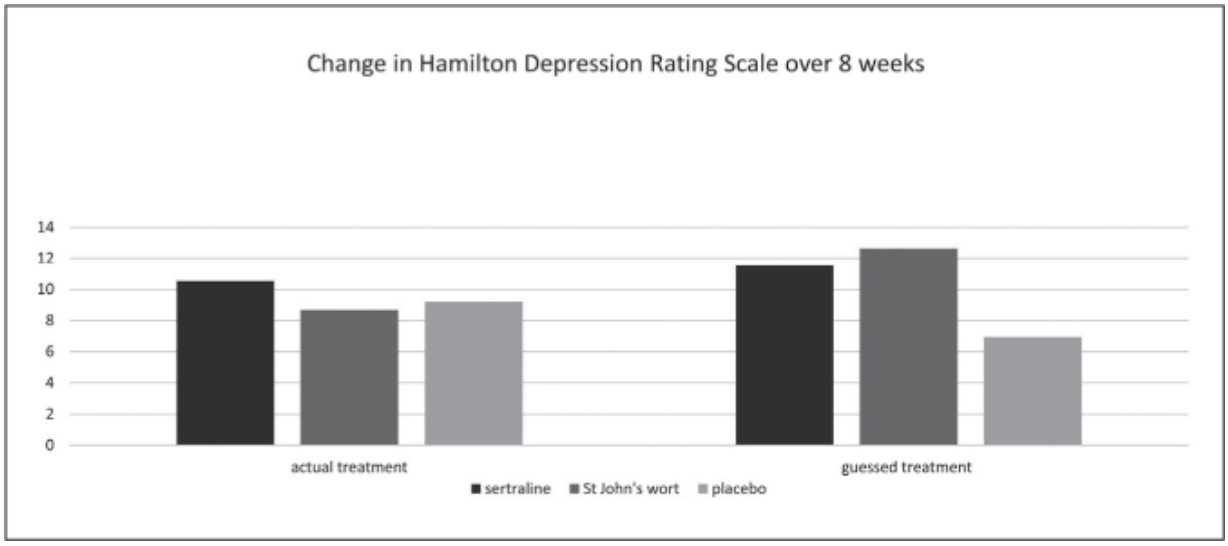
Thomson found these trials were less likely to detect an effect of antidepressants compared to trials using regular, inactive placebos. The

paper made such a strong impression on me that as soon as I managed to obtain a post where I could do my own research, I updated Thomson's review and confirmed that there is little, if any, difference between antidepressants and an active placebo.[14]

Since the 1970s, no one has paid much attention to this problem and there have been no further antidepressant trials using an active placebo. Yet, a recent study clearly demonstrates the power of people's beliefs about medication. Psychologists in Sweden enrolled forty-seven people who had been diagnosed with social anxiety disorder and gave them all the SSRI escitalopram, which is thought to be an effective treatment for anxiety.

Only half the participants were told the drug was escitalopram, however. The other half were told they were receiving a placebo. After nine weeks of treatment, the people who were told they were receiving the drug had a 51 per cent reduction in their social anxiety symptoms and those who were told they were being given the placebo only showed a 26 per cent reduction, a large and statistically significant (non-random) difference.[15]

A clinical trial that compared the effects of the SSRI sertraline, St John's Wort (a plant extract, also known as hypericum, which is occasionally used as an antidepressant) and a placebo in people with depression also shows the influence of people's belief about what they are taking.[16] As the left-hand columns in Figure 2 show, the results of this trial revealed minimal differences (1 point or less on the Hamilton Depression Rating Scale) between the treatments in the degree of improvement in people's depression scores, and these differences were not statistically significant (they could have occurred by chance).



Effects of guesses on depression symptoms in a randomised trial of sertraline vs St John's Wort vs placebo (derived from data in Chen et al., 2011).

The columns on the right-hand side, however, show that when the data were analysed according to what people *guessed* they were taking, rather than what they were *actually* taking, there was a more substantial and statistically significant difference. People who guessed they were taking sertraline improved by 5 points more than people who guessed they were taking the placebo, and the difference between those who guessed they were taking St John's Wort and those who guessed they were taking the placebo was almost 6 points.

Participants in this trial didn't guess what they were taking more accurately than would be predicted by chance, maybe because the trial involved two different active drugs, making guessing quite complicated. Therefore, people's guesses didn't influence the ultimate results of the trial, which was negative. It clearly demonstrates, however, that whether people think they are taking a real antidepressant or a placebo has a significant influence on their mood, regardless of what they are actually taking.

In contrast, in most trials in which people are asked to guess the identity of their tablets, they do guess better than chance.[17] In a trial of fluoxetine (Prozac) and placebo for the treatment of people with alcohol problems, for example, 80 per cent of those assigned to fluoxetine correctly guessed they

had been allocated the active drug, whereas only 45 per cent of the placebo group guessed (incorrectly) that they were taking fluoxetine.[18] When this occurs, the trial is not truly double-blind and, as a consequence, the antidepressant may exert an 'amplified placebo effect', which makes it look more effective than it really is.

The TADS study

Data from another study illustrate how this can happen. The Treatment for Adolescents with Depression Study, known as TADS, was conducted in the early 2000s in the USA and has been influential in producing a consensus that fluoxetine is an effective antidepressant for children and young people. The study had a complex design that involved comparing fluoxetine with a placebo, with some participants also having cognitive behavioural therapy and one group only having the therapy. The main paper, published in 2004, reported that adolescents treated with a combination of fluoxetine and cognitive behavioural therapy showed the greatest improvement in their depression scores, followed by those taking fluoxetine alone, followed by those assigned to a placebo or cognitive behavioural therapy alone.[19]

Years later, Professor Jon Jureidini of the University of Adelaide, Australia, obtained the data from the trial and led a group of researchers, including myself, in looking at the effects of people's guesses about the identity of their tablets, which had not been reported previously. Like the study of sertraline and St John's Wort, we found that what the adolescents guessed they were taking when they were asked six weeks into the study strongly predicted their mood ratings at the end of the study, six weeks later (the double-blind part of the trial lasted twelve weeks in all). But unlike in the previous study, everyone involved in the TADS trial – including the adolescent participants, their parents, their doctors and the researchers – could guess the identity of the tablets slightly better than expected. Instead

of being correct 50 per cent of the time, as would be predicted by chance, they were correct between 60 and 62 per cent of the time.

Our analysis was unique in that we were able to use statistical techniques to explore how people's guesses interacted with the effects of the drug. This showed that when we accounted for the effects of people's guesses, the actual nature of the medication (whether it was fluoxetine or placebo) had no impact on depression scores. However, when we removed the effect of guessing from the analysis, the apparent effect of the medication increased, and people assigned to fluoxetine showed a marginal statistical advantage over those allocated to the placebo.[20]

So, we showed that if you don't account for the effects of people's guesses when you conduct an antidepressant trial, you can end up with a spuriously positive result – the drug appears to be more effective than the placebo, when it isn't. This is the amplified placebo effect. If the TADS trial had involved an antidepressant with more noticeable effects than fluoxetine (which has relatively few side-effects), people's guesses might have been more accurate. In that case, the difference between the antidepressant and the placebo would likely have looked larger.

While most advocates of antidepressants ignore the issue of amplified placebo effects, some have suggested it is not relevant because people's guesses may be based on the benefits they get from the drug, rather than on its side-effects.[21] However, the TADS study showed that guesses could not be due to a therapeutic effect of the drug because there was none.

Effects of other substances in depression

Now we can see why it is relevant that people who enrol in randomised trials want to get the active drug and are highly invested in working out what sort of tablet they have been allocated to, as I observed in the alcohol treatment trial I was involved in long ago. People's beliefs about the likely effectiveness of a drug influence their ultimate outcome substantially, and if

they can guess whether or not they are taking the real drug, those beliefs can influence the results of the trial.

This explanation for antidepressants' effects makes sense of some curious facts about antidepressants. It explains why they are such a disparate collection of chemicals, with little in common except that they affect the brain in some way, and why numerous substances that are not normally thought of as antidepressants have been found to produce similar effects to antidepressants on symptoms of depression. These include benzodiazepines, stimulants, amphetamine, methylphenidate (Ritalin), opioids, buspirone (the 1980s anti-anxiety drug), many antipsychotics (including chlorpromazine) and 'purple hearts', as found by researchers back in the 1960s.[22]

It also makes sense of the most contradictory finding of all: a drug that has the opposite effect of the SSRIs, a serotonin reuptake *enhancer*, has been licensed and is used as an antidepressant. It is called tianeptine, and it is prescribed in some European and South American countries. Like other antidepressants, tianeptine has gone through clinical trials and been shown to be marginally better than a placebo.[23] What all these chemicals have in common is that they produce noticeable physical and mental changes or side-effects. So, the fact that they all produce much the same effect in depression suggests they are working through an amplified placebo effect.

Anxiety

Research on anxiety is similar to research on depression. There are fewer trials but, overall, they reveal that antidepressants improve anxiety symptoms a little better than a placebo. In a recent meta-analysis of trials involving people with general anxiety, antidepressants outperformed placebo by only 2–3 points on an anxiety symptom scale that has a maximum score of 56.[24] In obsessive compulsive disorder, usually classified

as an anxiety disorder, antidepressants show an advantage of 3 points on a 40-point scale.[25] [26]

There is no comparable research on the meaning of these differences in anxiety or obsessive compulsive disorder, but they don't look impressive. Moreover, people with anxiety are just as likely to be able to guess the identity of their tablets as people with depression, and to be influenced by positive expectations of treatment, as we saw in the experiment with people with social anxiety disorder. So, the evidence for the effects of antidepressants in anxiety are the same as for depression. Placebo-controlled trials reveal small differences that are unlikely to be relevant and may represent amplified placebo effects.

Antidepressant development

At this point, you might be wondering how particular drugs are selected as antidepressants, given what I have said about their variability. Surely, you might suppose, there is a scientific process of drug development that identifies drugs with antidepressant action on a rational basis?

I thought this was the case, too, until I started doing research for my first book. During my investigations, I came across a scientific review of antidepressant effects in 'animal models' of depression.[27] 'Animal models' are used to test drugs for antidepressant effects and involve inducing depression-like states in animals, such as rats and mice. The best known is called the Forced Swim Test, which involves making the animals swim in a tank they can't get out of and measuring how long it takes them to give up trying (sounds cruel, I know). Drugs that keep them swimming for longer are interpreted as having antidepressant properties.

I had assumed that antidepressants are identified using these models, and then tested in humans and brought to market. This is what is claimed by people who do these studies.[28] But it turns out animal model experiments are highly unreliable: tests done in different laboratories' yield different

results. Antidepressants don't always show the predicted effects and drugs that are not considered antidepressants, such as amphetamine, often show what are referred to as 'false-positive' effects (because they are not deemed to be antidepressants, they are not meant to show positive effects). Yet, given the stimulant properties of amphetamine, it is not surprising that it keeps animals swimming for longer.

Putting aside the obvious problem that animal models of complex human emotions are flawed to begin with, this suggests there is no systematic scientific process behind the identification of drugs that are deemed to be antidepressants. As you may have realised already, what gets branded as an antidepressant is determined instead by a combination of the latest vogue in neurotransmitter theory and what drug companies think will sell best at a particular time.

Other outcomes

I have described how antidepressants have small and irrelevant effects on symptoms of depression, but what do we know about how they affect people's actual lives? Do they help people to manage their day-to-day responsibilities better and enable them to enjoy their social relationships and leisure activities? There is little evidence on these questions for the simple reason that clinical trials focus primarily on depression symptoms. Sometimes, what is called 'quality of life' is measured using questionnaires, but these overlap considerably with depression questionnaires; their results are not consistently reported and they are also susceptible to amplified placebo effects, just like depression questionnaires. There is almost no data from clinical trials on such important things as whether antidepressants help people get back to work, how productive people are, how reliant they are on health services, or how people's intimate relationships, social lives and recreational activities are affected by treatment.

Conclusions

Recommendations for the use of antidepressants by authorities such as NICE (the National Institute for Health and Care Excellence) are based on the placebo-controlled trials I have described.[29] These trials set out to measure depression as if it were high blood pressure, disregarding the obvious fact that human emotions are not readily amenable to being quantified. They last a mere few weeks in most cases, and rarely provide objective data on how people are managing their lives. It is doubtful that antidepressant trials tell us much at all, therefore, but if we accept these trials on their own terms, they reveal that antidepressants are barely better than a placebo, and the small difference detected is likely to be accounted for by amplified placebo effects.

All this begs the question: how has this meagre evidence been transformed into the much-trumpeted idea that ‘antidepressants work’?