Joe Coyne
12/12/24

Clinical Trials Final Project

Maintenance or Discontinuation of Antidepressants in Primary Care

The goal of this study (link can be found in the Appendix) was to assess the effects of maintenance antidepressant therapy, as compared with discontinuation of treatment, in primary care patients who had been taking antidepressants for more than 9 months and felt well enough to consider stopping their medication. This study was conducted using the randomized Antidepressants to Prevent Relapse in Depression (ANTLER) trial.

The background as to why this study was run surrounds the fact that patients diagnosed with depression who are treated in primary care practices may receive antidepressants for prolonged periods. However, data is limited to the effects of maintaining or discontinuing antidepressant therapy in this setting. There have been several systematic reviews of studies that show a higher rate of relapse in depression among patients who discontinue antidepressant therapy than among those who continue to receive such therapy. However, these studies have had several limitations, including poor study setup and small sample sizes, limiting the ability to draw firm conclusions. This study was a multicentered one, recruiting patients from 150 general practices across four sites in England, namely London UCL, University of Southampton, University of Bristol, and Hull York Medical School.

The patients in this study were patients with a history of at least two depressive episodes or who had been taking antidepressants for 2 years or longer and felt well enough to consider stopping antidepressants. The study enrolled patients who were receiving conventional doses of the three most commonly prescribed antidepressants in the UK (citalopram, sertraline, and fluoxetine). Mirtazapine was also included among the trial drugs because of its increasing popularity and use in the United Kingdom. However, patients taking escitalopram, paroxetine, or venlafaxine were excluded. Escitalopram is not widely used in UK primary care, paroxetine prescription rates are dropping and discontinuation can lead to marked withdrawal symptoms, and

discontinuation of venlafaxine also causes withdrawal symptoms and most clinical guidelines recommend it as a second-line treatment.

Among the other eligibility requirements, patients had to be between the ages of 18 and 74 years and had to be receiving and adhering to a daily regimen of the following: 20 mg of citalopram, 100 mg of sertraline, 20 mg of fluoxetine, or 30 mg of mirtazapine for at least 9 months. Patients also had to have recovered from their most recent depressive episode and have felt well enough to consider stopping antidepressants. The main exclusion criterion in the study was current depression at the time of trial entry. Current depression in this study is defined using the criteria of the International Classification of Diseases, version 10 (ICD-10). This was determined by patients completing the original version of the Clinical Interview Schedule-Revised (CIS-R), which asks about depressive symptoms during the past week and determines whether the symptoms indicate a diagnosis meeting the ICD-10 criteria for depressive episodes.

The treatments administered during the study were either a maintenance/continuation of the patient's current antidepressant therapy or a taper and discontinuation of said therapy with the use of a matching placebo. This split was denoted in the study by creating two groups of patients: the maintenance group and the discontinuation group. In the maintenance group, the patients simply received their usual antidepressants at their usual doses. In the discontinuation group, patients taking citalopram, sertraline, or mirtazapine at baseline followed the following structure: taking half of their regular dose for the 1st month of the trial, taking the previously described half-dose and placebo on alternating days in the 2nd month of the trial, and finally taking the placebo only in the 3rd month and beyond. In the discontinuation group, patients taking fluoxetine at baseline followed the following structure: taking 20 mg of fluoxetine and the placebo on alternating days in the 1st month (fluoxetine was not available in a 10-mg capsule, the proper half dose at the time of enrollment), and taking the placebo only in the 2nd month and beyond (since fluoxetine has a long half-life). The physical trial medications contained antidepressants at full or half doses or a lactose pill representing the placebo. These medications were in lactose film-coated, over-encapsulated capsules, all manufactured by Capsugel and B&C Group. All the capsules had an identical opaque appearance and were provided in identical bottles.

This study is a parallel design study, since each patient either is part of the maintenance group or the discontinuation group (the two treatments) and is followed for 52 weeks until the end of the study. The study was a randomized, double-blind trial, where the trial-group assignments were provided to pharmacy staff members, who then sent masked trial antidepressants or placebo to the primary care practice for distribution to patients or directly to the patient's home. Both the primary care practice and the patients are blinded in this study. The comparator used in this study was a placebo, in this case a lactose pill. Using a computerized system and minimization algorithm that included site, medication, and median CIS-R score, patients who had received citalopram, fluoxetine, sertraline, or mirtazapine were randomly assigned in a 1:1 ratio to maintain their current antidepressant therapy (maintenance group) or to taper and discontinue such therapy with the use of matching placebo (discontinuation group). The trial was performed over a period of 52 weeks, with follow-ups on the patients at 6, 12, 26, 39, and 52 weeks. Data was collected through mailed questionnaires send to patients at 6 weeks, and through face-to-face interviews conducted at baseline and again at 12, 26, 39, and 52 weeks.

The primary hypothesis of this study was to assess the effects of maintenance antidepressant therapy (maintaining the current therapy), as compared with discontinuation of treatment, in primary care patients who had been taking antidepressants for more than 9 months and felt well enough to consider stopping their medication. The secondary hypotheses were to assess the effects of maintenance antidepressant therapy on depressive and anxiety symptoms, physical and withdrawal symptoms, quality of life, time to stopping an antidepressant or placebo, and global mood ratings. The primary endpoint was the first relapse of depression during the 52-week follow-up period. This outcome was defined as a new episode of depression, as determined by components of a modified retrospective CIS-R (rCIS-R) that was adapted for the purpose of this trial. As opposed to the original rCIS-R, this one inquired specifically about the patient's experience during the previous 12 weeks. This study's definition for relapse of depression was a "yes" answer to either of the following two questions: "Have you ever had a spell of feeling sad, miserable, or depressed?"; "Have you been unable to enjoy or take an interest in things as much as you usually do?". To meet the outcome of a depressive episode, patients also had to report that at least one of the previous responses had lasted for 2 weeks or more and to describe the occurrence of at least one of the following symptoms: depressive thoughts, fatigue, loss of concentration, or sleep disturbance.

The secondary hypothesis of depressive symptoms was measured using the Patient Health Questionnaire 9-item version (PHQ-9). The secondary hypothesis of general anxiety symptoms was measured using the Generalized Anxiety Disorder Assessment 7-item version (GAD-7). Physical symptoms that were potentially side effects of antidepressant therapy were reviewed using a modified 13-item Toronto Side Effect Scale. The frequency of new or worsened drug-withdrawal symptoms was evaluated on a modified 14-item Discontinuation-Emergent Signs and Symptoms (DESS) checklist. Quality-of-life scores for physical and mental health categories were calculated using the 12-Item Short-Form Health Survey (SF-12). The time to stop an antidepressant or placebo was calculated with the interval between the date of initiation of an antidepressant or placebo and the stopping data. Finally, the patient's reported global rating of mood was calculated (feeling worse was denoted "grade 1" and feeling the same or better was denoted "grade 0"). All outcomes were assessed at every follow-up except for scores on the rCIS-R, SF-12, and adherence scales, which were obtained at every follow-up except at 6 weeks.

A 90% power was used in this study to detect a hazard ratio of 1.92 for the primary outcome in the discontinuation group, assuming a 20% attrition/dropout rate. The hazard ratio is calculated on the basis of an estimated relapse frequency of 20% in the maintenance group and 35% in the discontinuation group at a two-sided alpha level of 0.05. The study determined that a total of 479 patients would be needed to reach this power. In order to recreate this sample size calculation in SAS using PROC POWER under the COX REG option, we would need to know the desired power, the alpha level, and the hazard ratio, all of which is given in the paper. But we would also need to know the estimated standard deviations of the two groups in terms of the primary outcome, risk of experiencing a relapse of depression within the 52 weeks of the study, but this is not given. In the actual study, a total of 478 patients were enrolled and participated in the study, which is only 1 less than the calculated amount needed to reach a 90% power, so this doesn't seem like that big of an issue going forward in the study.

In terms of the final results and conclusions of the study, after the 52 week trial had concluded, relapse in depression (the primary outcome) was noted in 92 of 238 patients (39%) in the maintenance group and in 135 of 240 (56%) in the discontinuation group. This calculates to a hazard ratio of 2.06, meaning that having a relapse in depression is 2.06 times more likely to occur in the discontinuation group compared to the maintenance group.

The secondary outcomes, like depressive and anxiety symptoms, physical and withdrawal symptoms, quality of life, time to stopping an antidepressant or placebo, and global mood ratings, were generally in the same direction as the primary outcome. Patients in the discontinuation group had more symptoms of depression, anxiety, and withdrawal than those in the maintenance group. After 12 weeks, secondary outcomes were generally in the same direction as the primary outcome, except for the scores on the SF-12 physical health component and the Toronto Side Effect Scale. At other time points, effect estimates were also generally in the same direction as those for the primary outcome, although the confidence intervals in several categories crossed the null cutoff (encompassed the point 0.0 in the interval). At the 12 week follow-up, the mean scores for side effects as assessed on the Toronto scale were 4.2 +/- 2.9 in the maintenance group and 4.6 +/- 3.0 in the discontinuation group, where higher scores indicate more severe symptoms. This is an estimated difference of 0.7 points, where the discontinuation group had worse results. At the 12 week follow-up, the mean scores for withdrawal symptoms on the modified DESS scale were 1.3 +/- 2.4 in the maintenance group and 3.1 +/- 3.5 in the discontinuation group, where higher scores indicate more severe symptoms. This is an estimated difference of 1.9 points, where the discontinuation group had worse results. And finally, at the 12 week follow-up, the mean scores for mental health-related quality of life on the SF-12 scale were 46 +/- 10 in the maintenance group and 41 +/- 11 in the discontinuation group, where higher scores indicate better quality of life. This is an estimated difference of -4.9 points, where the maintenance group had better results.

Some other results from the trials are also listed. A greater percentage of patients in the discontinuation group stopped taking the trial medication before the end of the trial as compared to patients in the maintenance group (48% compared to 30%). This calculates to a hazard ratio of 2.28 with a 95% confidence interval of 1.68 to 3.08. Of the patients who did stop their trial medication, the percentage who returned to the use of an antidepressant prescribed by their primary care doctor was 20% in the maintenance group and 39% in the discontinuation group. During the trial, 157 of 225 patients (70%) in the maintenance group adhered to the trial regimen, compared to 119 of 230 patients (52%) in the discontinuation group. And finally, at 12 weeks, patients who reported feeling worse than they had felt at 6 weeks included 48 of 228 patients (21%) in the maintenance group and 94 of 216 patients (44%) in the discontinuation group. The final conclusion of the study was that among patients in primary care practices who felt well enough to discontinue

antidepressant therapy, those who were assigned to stop their medication had a higher risk of relapse of depression by 52 weeks than those who were assigned to maintain their current therapy.

Overall, there weren't too many issues with the paper. Most of the analyses were well-run, and the conclusions were appropriate. However, there were a few areas where either more data was needed to better replicate the calculations or there were some slight inconsistencies with data reporting. Figure 2 in the Appendix is a replication of the Kaplan-Meier Estimates of the Primary Outcome from the paper. In order to have a more accurate representation of these two curves and estimates, we would need the individual data on each patient and when/if they had a relapse in depression to more accurately denote a change in the graph and slopes. Having the individual patient data is also necessary to calculate the final survival time function value. In my recreation, the number of relapses in a given time period (d_i) was calculated by taking the difference in the number at risk from the current time and the most recent time and subtracting the number of patients censored between these two times. While this gives a general idea of how the survival graphs should look, having the individual data would provide a much more accurate and informative graph.

A simple replication of Table 2, the mean differences in scores regarding the primary outcome and some secondary outcomes, is also shown in the Appendix. This is another instance where having the individual patient data would be extremely helpful in recreating analyses. A simple PROC TTEST for difference in means can be run in SAS to easily replicate Table 2, but the individual data points from the trial are needed, and the paper only gives some of the summary statistics (namely mean, standard deviation, and sample size).

One area where there are some data reporting inconsistencies is on the Research Summary page. When comparing the scores for secondary outcomes, the values used in the charts are from the readings taken at 26 weeks. However, the only way you would know this is if you cross-validated this Research Summary Table with Table 2 in the main paper, which lists the secondary outcome scores at 12 weeks, 26 weeks, 39 weeks, and 52 weeks. This can lead the reader to make incorrect or misleading conclusions. For instance, when looking at the scores from the modified Toronto Side Effect Scale, the maintenance group has a lower mean score of 4.2 compared to the

discontinuation group, whose mean score is 4.6. But at 26 weeks, the maintenance group now has a greater mean score compared to the discontinuation group (4.0 compared to 3.9).

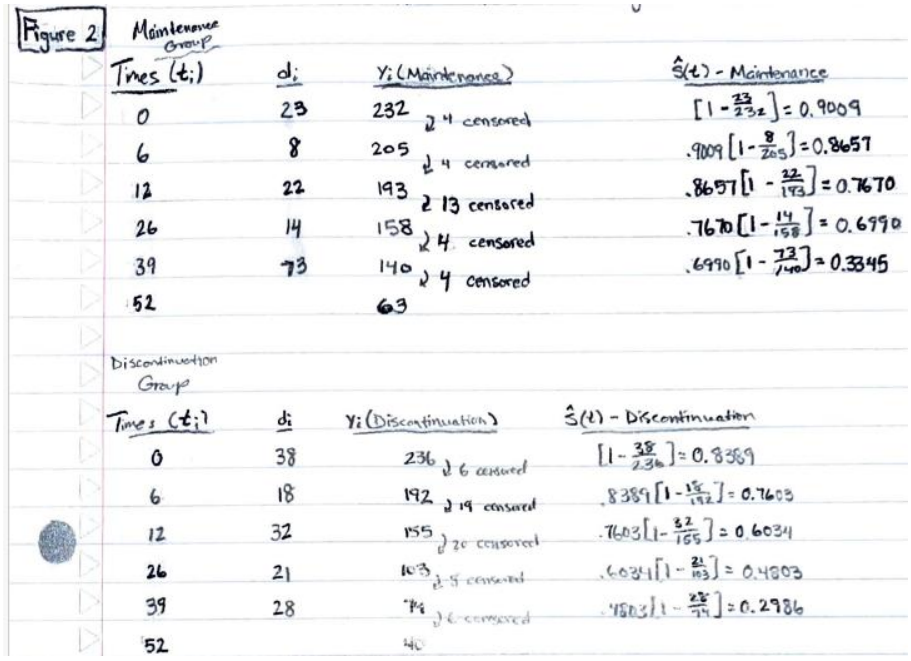
The two main analyses the paper used were the Cox Proportional Hazard Model, which the authors used in their primary analysis after adjustment for the baseline CIS-R depression score, and the Kaplan-Meier Analysis, which the authors used to determine proportionality and the predicted survival plot. The assumptions for a Cox Proportional Hazard Model are that there are independent observations, independent censoring, and that the survival curves for two different strata of a risk factor must have hazard functions that are proportional over time. The assumptions for a Kaplan-Meier analysis are that the censoring is non-informative, that survival times are independent, and that there are no competing risks, ensuring accurate survival estimates in the presence of censored data. The paper claims that all of these assumptions were met, and I have every right to believe they were.

One final area of the paper that gives me reason for pause is some of the demographic distributions of the patients. 73% of all the patients were women, compared to only 27% of men, which is concerning because this is a large difference and men and women may have different reactions to stopping depression medication/therapy. I would be more comfortable with the study and conclusions if there was close to an equal split of men and women in the study. Most of the patients were also older in age, with the mean age in the maintenance group being 54 and the mean age in the discontinuation group being 55. But the biggest area of concern lies in the fact that a large majority of the patients in both groups self-reported themselves as white (93% in the maintenance group and 97% in the discontinuation group). The authors mention that the trial population lacked ethnic diversity, so there wasn't really any way to combat this race disparity. In the discussion section, this is mentioned along with the fact that all the patients were being treated in the U.K. health system, so the results cannot be generalized to non-White patients and to other health systems.

In general, this paper does a very good and thorough job of detailing the trial and accompanying analyses outside of a few data discrepancy and clarification issues. The results appear to be correct and the conclusions drawn are appropriate and easy to understand.

Appendix

Figure 2 - Kaplan–Meier Estimates of the Primary Outcome



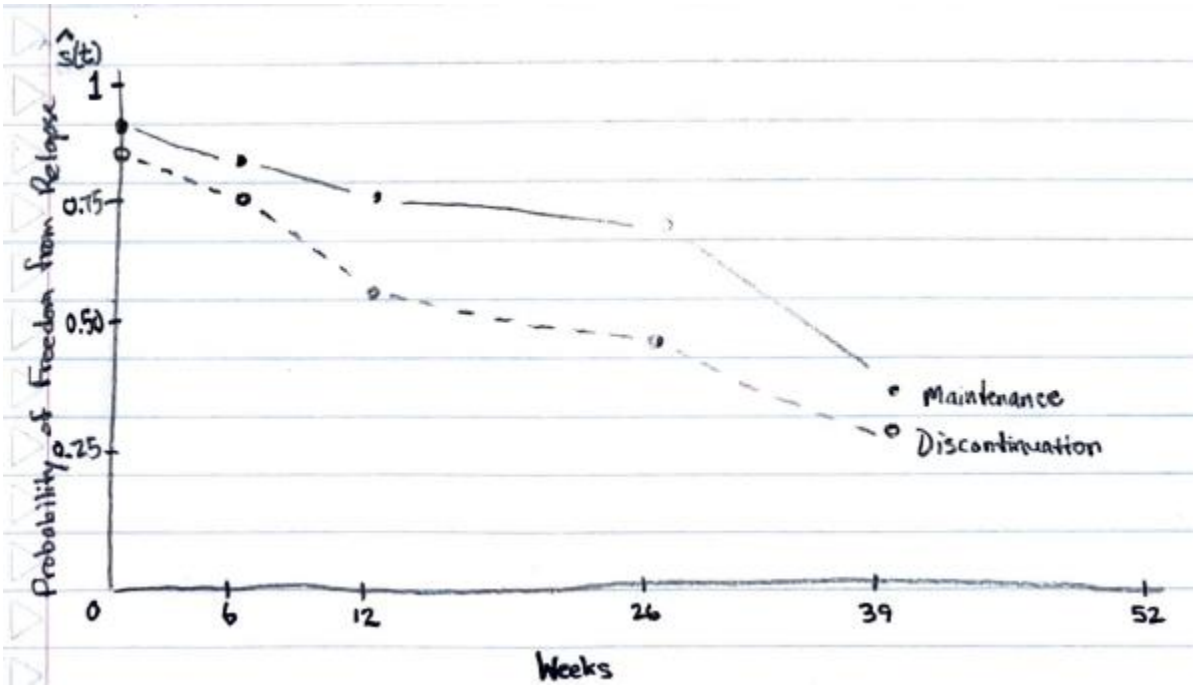


Table 2 - Difference in Means for Primary and Secondary Outcomes

TABLE 2

	Maintenance Group (N = 238)		Discontinuation Group (N = 240)	
Primary Outcome				
Relapse of depression - no. (%)	92 (39)		135(56)	
Secondary Outcomes				
Score on Patient Health Questionnaire 9-item version				
12 wk	4.1	+/- 3.8	6.3	+/- 5.1
26 wk	4.2	+/- 3.7	5	+/- 4.6
39 wk	3.8	+/- 3.9	4.4	+/- 4.2
52 wk	3.7	+/- 3.7	4	+/- 4.5
Score on Generalized Anxiety Disorder 7-item version				
12 wk	3.1	+/- 3.3	5.3	+/- 4.6
26 wk	3.4	+/- 3.8	4.1	+/- 4.4
39 wk	2.9	+/- 3.5	3.8	+/- 4.1
52 wk	3	+/- 3.7	3.1	+/- 3.0
Score on Modified Toronto Side Effect Scale				
12 wk	4.2	+/- 2.9	4.6	+/- 3.0
26 wk	4	+/- 2.6	3.9	+/- 2.8
39 wk	3.8	+/- 2.5	3.7	+/- 2.6
52 wk	3.7	+/- 2.6	3.5	+/- 2.8

TABLE 2

	ffect Size or Difference (95% CI)	***Actual Effect Size or Difference***
Primary Outcome		
Relapse of depression - no. (%)	E4/C4" = 1.467	***Hazard Ratio, 2.06 (1.56 to 2.70)***
Secondary Outcomes		
Score on Patient Health Questionnaire 9-item version		
12 wk	2.2	***2.2 (1.5 to 2.8)***
26 wk	0.8	***0.7 (0.0 to 1.4)***
39 wk	0.6	***0.6 (-0.1 to 1.2)***
52 wk	0.3	***0.4 (-0.3 to 1.1)***
Score on Generalized Anxiety Disorder 7-item version		
12 wk	2.2	***2.4 (1.8 to 3.0)***
26 wk	0.7	***0.8 (0.1 to 1.4)***
39 wk	0.9	***1.0 (0.4 to 1.6)***
52 wk	0.1	***0.3 (-0.4 to 0.9)***
Score on Modified Toronto Side Effect Scale		
12 wk	0.4	***0.7 (0.3 to 1.1)***
26 wk	-0.1	***0.2 (-0.3 to 0.7)***
39 wk	-0.1	***0.2 (-0.4 to 0.5)***
52 wk	-0.2	***0.0 (-0.4 to 0.5)***

	Maintenance Group (N = 238)		Discontinuation Group (N = 240)	
No. of new or worsening symptoms on modified DESS				
12 wk	1.3	+/- 2.4	3.1	+/- 3.5
26 wk	1.4	+/- 2.3	1.9	+/- 2.9
39 wk	0.8	+/- 1.6	1.7	+/- 2.7
52 wk	0.8	+/- 1.8	1.1	+/- 2.5
Score on physical component of 12-Item Short-Form Health Survey				
12 wk	48	+/- 10	50	+/- 9
26 wk	48	+/- 10	49	+/- 10
39 wk	48	+/- 11	51	+/- 10
52 wk	49	+/- 10	49	+/- 11
Score on mental component of 12-Item Short-Form Health Survey				
12 wk	46	+/- 10	41	+/- 11
26 wk	46	+/- 11	44	+/- 11
39 wk	48	+/- 10	45	+/- 11
52 wk	47	+/- 10	46	+/- 11

	Effect Size or Difference (95% CI)	***Actual Effect Size or Difference***
No. of new or worsening symptoms on modified DESS		
12 wk	1.8	***1.9 (1.5 to 2.3)***
26 wk	0.5	***0.5 (0.1 to 0.9)***
39 wk	0.9	***0.9 (0.6 to 1.3)***
52 wk	0.3	***0.3 (-0.0 to 0.6)***
Score on physical component of 12-Item Short-Form Health Survey		
12 wk	2	***0.4 (-0.9 to 1.8)***
26 wk	1	***0.2 (-1.3 to 1.6)***
39 wk	3	***1.5 (-0.1 to 3.0)***
52 wk	0	***-0.6 (-2.1 to 0.9)***
Score on mental component of 12-Item Short-Form Health Survey		
12 wk	-5	***-4.9 (-6.4 to -3.3)***
26 wk	-2	***-2.6 (-4.4 to -0.8)***
39 wk	-3	***-3.1 (-4.8 to -1.3)***
52 wk	-1	***-1.6 (-3.4 to 0.2)***

Sources:

https://www.nejm.org/doi/full/10.1056/NEJMoa2106356