Subject: BMJ - Decision on Manuscript ID BMJ-2019-050262

Body: 09-Jun-2019

Dear Dr. Bråten,

Manuscript ID BMJ-2019-050262 entitled "Efficacy of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM study): a double-blind, randomised, placebo-controlled, multicentre trial"

Thank you for sending us your paper. We sent it for external peer review and discussed it at our manuscript committee meeting with editors and our statistical consultant in attendance. We are very interested in publishing the paper, but would like you to respond to the comments of reviewers and editors, as explained below in the report from the manuscript meeting. We look forward to seeing a revised version of the paper and reaching a final decision.

Please remember that the author list and order were finalised upon initial submission, and reviewers and editors judged the paper in light of this information, particularly regarding any competing interests. If authors are later added to a paper this process is subverted. In that case, we reserve the right to rescind any previous decision or return the paper to the review process. Please also remember that we reserve the right to require formation of an authorship group when there are a large number of authors.

When you return your revised manuscript, please note that The BMJ requires an ORCID iD for corresponding authors of all research articles. If you do not have an ORCID iD, registration is free and takes a matter of seconds.

Very truly yours,

Elizabeth Loder, MD, MPH

*** PLEASE NOTE: This is a two-step process. After clicking on the link, you will be directed to a webpage to confirm. ***

https://mc.manuscriptcentral.com/bmj?URL_MASK=1221506a51ed4f1392ca25dd17cca16f

Report from The BMJ's manuscript committee meeting

These comments are an attempt to summarise the discussions at the manuscript meeting. They are not an exact transcript.

Present: Elizabeth Loder (chair); Gary Collins (statistician); Tim Feeney; David Ludwig; John Fletcher; Tiago Villanueva; Daoxin Yin

Decision: Request revisions before final decision

* We thought the protocol was excellent, and that some information from it should be in the main paper. For example, we would like more information about the sample size calculation in the main body of the paper.

* We discussed reviewer concerns about the effect size and concluded that this study was set up to evaluate whether the effect identified in the original study (which was not, we believe, an ITT analysis) was believable. If that is the aim of the current study, then it seems unfair to criticize it for not excluding smaller effects. We agree with how you have expressed the results and also on how you calculated the sample size. The previous trial had suggested a very large effect of antibiotics: 8 points on the RMDQ and many thought this just wasn't believable. In that context you set out to see if the effect of antibiotic really

could be that big. You were not looking for any old effect but for one of comparable magnitude to the previous study, so RMDQ of 4 is conservative for this purpose. 160 patients in a back pain RCTs is a respectable sized trial and the result clearly rules out the possibility of a difference of 8 in the RMDQ. The point estimate of 1.6 is smaller than most estimates of what would matter clinically. It is true that the upper bound of the CI includes a difference of 4, which some might argue is clinically meaningful, but this must be interpreted in light of the fact that it is an extreme value. You might acknowledge that there is this possibility but that it is quite unlikely. You should also address the matter of the I and II subgroups.

We think the overall message, however, is as you have said: you couldn't replicate the previous study and so it is probably not correct. We encourage you to make slightly stronger reference to this in the introduction (e.g. "we set out to see if we could replicate the findings...") and conclusions (e.g. "we were not able to replicate the findings...").

* You might also explain in even more detail why you set the bar where you did. This is explained in the protocol and is one of the things in the protocol that probably belongs in the paper.

* Please look at instructions for authors on bmj.com to see our requirements for data sharing statements. Your current statement does not fit our requirements.

* Trials are not performed according to CONSORT guidelines. CONSORT is a guideline for reporting, not conducting, trials.

* Please make sure that you report outcomes as specified in the trial registry. If you report any outcomes not listed in the trial registry, please clearly label them as post-hoc outcomes. (For example, Outcomes at 3 months were not specified in the registry as far as we can see.) The paper itself is brief on outcomes – referring the reader to supplementary material – we would prefer all outcomes (and their timing) to be in the main body of the text.

* "The allocation sequence was concealed" – this is vague. Can you give more details?

* We agreed with reviewer focus on blinding. If people in the placebo group really did understand they were getting a placebo then all of the findings could be attributed to that, so maybe sensitivity analyses would be helpful to see if the people who perceived they were in the placebo group were scoring symptoms differently.

* We were impressed with the extent of patient involvement in the trial.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how and where you have dealt with them in the paper.

Comments from Reviewers

Reviewer: 1

Comments:
MAJOR COMMENTS
This is an important trial that sets up to replicate the Albert trial that concluded that antibiotics were a highly effective treatment for a sub-type of chronic LBP associated with Modic changes. At the time the Albert trial created great controversy because the theory underlying the treatment represented an entirely new way of viewing chronic LBP and the results achieved were much larger than typically seen with treatments for chronic LBP.

The current trial was prospectively registered and the manuscript closely follows the protocol and statistical analysis plan. The trial was well conducted and has excellent follow-up for the primary outcome. The analysis and reporting are both excellent and I only have a minor concern re reporting of treatment compliance which I will cover below. My one major concern is that I am not convinced that the interpretation of the trial result is appropriate.

Table 2 shows favourable treatment effects on 5 of the 6 outcomes (including the primary outcome) at 12 months follow-up. The trial also shows larger treatment effects in the subgroup with Modic type I changes (a pre-specified sub-group analysis with biological plausibility). Yet the authors conclude that the treatment should not be used. I do not think it is anywhere near as clear cut as this. The authors dismiss the effect on the primary outcome as too small to be clinically worthwhile; but the context is that there are few other options for this subset of patients, and I think that needs to be borne in mind. The effect is also larger than the effect in the UK BEAM (Back Exercise And Manipulation) trial published in the BMJ in 2004. The BEAM authors took a different approach and rather than saying the treatments should not be used, they noted the small to moderate effect sizes. Both treatments are endorsed in the current NICE guideline for LBP.

I would encourage the authors to reconsider their portrayal of the result and rather than advocate that the treatment should not be used, take a more nuanced approach. It would be reasonable to note that the treatment is effective but it typically has modest effects that are somewhat larger in the subgroup with Modic type I changes. This is a perfect scenario where shared decision making could be used so that an appropriately informed patient could decide if the likely benefit is worth the likely harms and inconvenience.

Another perspective on judging if this treatment is worthwhile would be consideration of an economic evaluation. I note that this was planned but not included. A 3 month course of antibiotics is quite cheap compared to some of the other alternate treatments offered for chronic LBP e.g a pain management program would cost many thousands of dollars, spinal fusion is even more expensive. Having this economic evaluation as a companion paper would really aid interpretation of this important trial. I would prefer what was done previously with the UK BEAM trial with two papers published at the same time in BMJ.

I acknowledge that my suggestion to revise the discussion and include the economic evaluation as a companion paper may disappoint the authors but I really think that this would be a far better outcome and the set of papers would have a much greater impact.

MINOR COMMENT
COMPLIANCE: It is unclear how treatment compliance was measured and Figure 1 provides inadequate information to judge the number of patients who complied with 13 week regimen. Please provide more detail here.

Additional Questions:
Please enter your name: Chris Maher

Job Title: Professor

Institution: The University of Sydney

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: Yes

Reviewer: 2

Comments:
First of all I want to congratulate the authors with this relevant, well designed, well executed
and reported study.

The relevance for conducting this (replication) study is undisputed. The single previous trial
evaluating antibiotic treatment for this subgroup of chronic back pain patients (Albert et al,
2013) showed very large and significant effects compared to placebo. That study gave high
hopes and expectations, but there were also questions and doubts (could it really be true?).

The present Norwegian study is the first replication.

I have a few questions and remarks:

1. a main observation and at the same time point of discussion is the predefined
between-group difference of 4 points on the RMDQ. This is really a large difference, and if
present there will be no discussion that that indeed is a clinically relevant outcome. The
question, however, is whether smaller differences are not clinically relevant ? In their study
protocol the authors themselves state that "the minimal important difference in mean RMDQ
between groups is not clear"! Subsequently they state that a change in RMDQ of 2-3 in
individual patients...is unlikely to be important and may present measurement error.
Subsequently they choose 4 as their cut-off point. We do, however, studies on group-level
and calculate mean effects, partly to get rid of measurement-error in individual patients.
More importantly there is some consensus to consider smaller between-group effects on the
RMDQ as being clinically relevant. For example the highly influential RCTs evaluating the
efficacy of stratified care using the Start Back Tool in the UK (Hill et al, Lancet 2011) and in
the USA (Cherkin et al, J Gen Int Med, 2018) used a between group difference of 2.5 points
on the RMDQ as being clinically relevant. A consequence of this smaller difference is of
course that we need larger sample size in our studies to test if this difference is present or
not..

2. Somewhat related to 1 is the observation that two subgroups of patients were included
MC type I and MC type II. This was well planned and motivated, but the sample sizes to
investigate the effect of antibiotic treatments were small and also varied within the trial.
Especially the number of patients with MC type II is relatively small.

3. Regarding the Bang blinding index I was wondering if in the placebo group the number of
patients guessing right (indeed allocated to placebo) was equally distributed over the MC
type I and MC type II groups?

4. I noticed that in one of the responder analysis (>30% improvement from baseline) the results favored the antibiotic group. This may well be an artifact (due to multiple testing) and can as such be discussed, but at present is does not get much attention in the paper.

These were my main observation when reading the study. In the light of these I wondered whether the current conclusions were not too harsh? For MC type I a significant effect is found which in (many) other trials would have been considered as a clinically relevant difference. For MC type II indeed no effect was found (but this subgroup was not included in the first RCT on this topic + the sample size of this subgroup in the present study was relatively small)

All in all, a great study. The conclusion may need some consideration and nuance.

Additional Questions:
Please enter your name: Bart Koes

Job Title: Professor of General Practice

Institution: Erasmus MC

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may
in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way
gain or lose financially from the publication of this paper?: No

If you have any competing interests <A
HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/dec
laration-competing-interests'target='_new'> (please see BMJ policy) please declare
them here:

Reviewer: 3

Comments:
This study tested the efficacy of amoxicillin in low back pain associated with Modic changes. While antibiotic treatment is unusual in low back pain, the results of a previous positive study may encourage health care providers to offer this treatment, in the light of the limited treatment options. Low back pain with Modic changes is prevalent and difficult to treat, resulting in an enormous number of potential candidates for antibiotic treatment. In the light of the growing antibiotic resistance worldwide and predictions that antibiotic resistance may become a major source of morbidity and mortality, the use of antibiotics for low back pain

may have disastrous consequences on overall population health. This study is therefore very timely.

Specific comments
Why was lumbar disc herniation an inclusion criterion? Modic changes can cause low back pain independent of lumbar disc herniation and no rationale is presented why antibiotics should work on pain associated with lumbar disc herniation. Furthermore, lumbar disc herniation typically causes radicular pain and not primarily low back pain.

It is unclear whether patients were assessed for radicular pain and, if so, how many patients in the two groups had radicular pain.

What was the rationale for choosing amoxicillin rather than another antibiotic?

I have a major concern. There is no information on the treatments that patients underwent during the follow-up phase. If placebo worked less well than amoxicillin, patients in the placebo group may have had more intensive treatments, theoretically minimizing a potentially significant difference between the two groups.

Additional Questions:
Please enter your name: Michele Curatolo

Job Title: Prof.

Institution: University of Washington

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may
in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way
gain or lose financially from the publication of this paper?: No

If you have any competing interests <A
HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/dec
laration-competing-interests'target='_new'> (please see BMJ policy) please declare
them here:

Reviewer: 4

Comments:
This is a well-written manuscript on a well-performed trial. The methods and reporting are adequate and comply with the CONSORT standards. The results, interpretation and conclusions are credible and well presented.
I have two minor comments on the flowchart only:

- In the boxes 'x patients evaluated at 1-year follow-up', you report the number of patients with major protocol deviations. It is unclear what is meant by 'end of study before 3 months'?
- The numbers of treatment non-completion in the two groups are n=10 for amoxicillin and n=8 for placebo. In other boxes, the number of treatment non-completion are also reported. Here it says n=7 (amoxicillin) and n=8 (placebo). It's not clear why these numbers are different.

Additional Questions:
Please enter your name: Trudy Bekkering

Job Title: researcher

Institution: Academic Center of General Practice, KU Leuven

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here:

Reviewer: 5

Comments:
I found this article to be of great interest & would recommend it for publication.

- Well written. Easy to follow.
- Excellent research & analysis.
- A valuable contribution to existing research - particularly in regards to 'treament routes' & this not to go down!
- Relevant to current concerns about the over prescribing of antibiotics. Noting hat I had never previously come across such use of antibiotics for the condition specifically focused on in this research/article. Hi

Additional Questions:
Please enter your name: Robert J. Nisbet

Job Title: Expert by Experience

Reviewer: 6

Comments:
This is a well-designed clinical study. The authors' randomized, double-blind, controlled
clinical trial showed that antibiotics could not treat chronic low back pain with Modic
changes.
Since Stirling et al. reported the presence of low virulent bacteria in herniated intervertebral
discs, a large number of basic and clinical studies have reported very contradictory results.
Albert et al. reported shocking results in their randomized, double-blind, controlled clinical
trial. They treated a group of patients with low back pain and Modic changes with the
antibiotic amoxicillin, compared with the control group, and the RMDQ score decreased by
8.3 points.
Is intervertebral disc degeneration with Modic changes associated with bacterial infection? Is
it true or not? If true, it may radically change our understanding of degenerative disc
disease. If it is false, taking antibiotics in large quantities for a long time will bring great side
effects to patients. We, doctors and patients, may all thank the authors of this well-designed
clinical trial.

I have a few concerns about this manuscript:
1. Previous studies have suggested that disc degeneration with Modic type 1 changes is
associated with low virulent bacterial infections. Why should the low back pain patients with
Modic type 2 changes be observed as a group? After all, the long-term use of antibiotics can
have great side effects.
2. Amoxicillin has a characteristic odor that patients may feel and may have some effect on
double blindness. In addition, long-term amoxicillin treatment group patients will have
greater gastrointestinal reactions, especially experienced patients.
3. Which of the patients in the antibiotic treatment group and placebo control group used
NSAIDs? What is the impact on the results?

4. The discussion section needs to be strengthened. Authors should draw clear conclusions around the current debate.

Additional Questions:
Please enter your name: Baogan Peng

Job Title: Spinal surgeon, Professor, Chairman

Institution: General Hospital of Armed Police Force, Beijing, China

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may
in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way
gain or lose financially from the publication of this paper?: No

If you have any competing interests <A
HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/dec
laration-competing-interests'target='_new'> (please see BMJ policy) please declare
them here:

Reviewer: 7

Comments:
This is a well written manuscript presenting the results of a well designed RCT on the
efficacy of antibiotic treatment in patients with chronic low back pain (LBP) and Modic
changes. The RCT was prospectively registered and a sample size calculation was performed
with a between-group difference used also in the results interpretation. The risk of bias of
this RCT may be low as there was no selection bias, blinding of patients, health professionals
and assessors, no attrition bias, and no selective reporting. My main comments regard (the
lack of): a) monitoring compliance to treatment and co-interventions, b) the chosen cut-off
to determine a between-group clinically important difference. I also have other comments
for the editors and authors to take into account.

1) Introduction, second paragraph. I miss (part of) the rationale on why running a new RCT
on the efficacy of antibiotic treatment in patients with chronic LBP and Modic Changes. This
may seem obvious to the authors but not to the readers. Why were the conclusions of the
previously RCT "questioned"? How was its methodological quality/risk of bias? Is the AIM
study trying to replicate the previously published trial?
2) Introduction, last paragraph. The previous published RCT included only patients with type
I Modic Changes. Why is the AIM study including both type I and type II? I believe this
should be included in the introduction/rationale for running this new RCT.

3) Methods. I could not find any information on whether compliance in each RCT arm was evaluated and if there were any between-group differences. If no compliance was monitored, this should be added as a limitation of this study.

4) Methods. "Included patients were recommended not to start additional treatments for back pain but were allowed to continue ongoing therapy. They were encouraged not to use NSAIDs during the intervention period." Were co-interventions monitored during the trial? Since I could not find information in the results section, I would encourage the authors to report some information if co-interventions were monitored. If co-interventions were not monitored, this should probably be added as a limitation of this study.

5) Methods and results. "To confirm MCs seen on a clinical MRI available at screening, baseline MRI was performed at median 22 days before treatment (...)". I wonder whether a disk herniation (inclusion criterion) was confirmed for all the patients included in the trial as this information is not reported in the results.

6) Statistical analysis. The authors adopted multiple imputation to accound for missing data on the primary outcome and perform intention-to-treat analysis. However, to my knowledge, intention-to-treat analysis does not have to do with missing data, as patients who have missing data do not necessarily change the trial arm that they are randomized to (ITT principle). Therefore, I was wondering if this approach is correct and a statistical advisor may be better suited than me to evaluate this aspect.

7) Statistical analysis. The authors performed linear-mixed models which have the assumption that regression residuals should be normally distributed. This is not always the case when using ordinal outcomes such as pain intensity in patients with LBP, therefore I would encourage the authors to report if this basic assumption of linear regression was met for all the outcomes.

8) Statistical analysis, minimal important between-group difference. I checked in the protocol why the authors decided for a "clinically important between-group mean difference of 4 RMDQ points" and, since this is fundamental information for the interpretability of this RCT, I would report it in this manuscript as well. As claimed by the authors, there is indeed no evidence supporting the chosen 4-point between-group difference, although it is larger than the primary outcome measurement error which is useful information. Considering this RCT RMDQ baseline mean values, 4 points correspond to about a 30% extra-improvement in the antiobiotics group as compared to the placebo group. The only data we have on medications in patients with chronic LBP suggest that patients expect a 30% extra-improvement if taking NSAIDs as compared to no intervention (Ferreira 2013 J Clin Epidemiol). This data is very much in line with the chosen cut-off, but the medication is different (antibiotics vs NSAIDs) and the control group is different (placebo vs no intervention). It would be important to further discuss the chosen 4-point clinically important between-group difference in the discussion.

9) Exploratory subgroup analysis Modic Changes type I vs type II. The results show a (small?) effect for type I Modic Changes but not for type II. These results are more in line with the already published RCT, although the magnitude of effect is substantially different between the two RCTs. How could the differences in the two RCTs results be explained?

10) Adverse events. The number of patients experiencing an adverse event in the placebo group (34%) is rather high for a placebo intervention. This could also be mentioned and explained in the discussion.

11) Bang blinding index. The between-group difference on this index is remarkable and, in my opinion, may be the major threat to the validity of this RCT. Especially if this potential bias of performance is added to the potential bias of performance related to the lack of monitoring of treatment compliance and co-interventions (see my points 3 and 4). The authors acknowledge this as a limitation but I would argue that this may be the major limitation of this RCT, which may the major threat to the reliability of this RCT effect estimates.

11) Discussion, second limitation. Should "overpowered" be "underpowered"?

Additional Questions:
Please enter your name: Alessandro Chiarotto

Job Title: Post-doctoral researcher

Institution: Department of General Practice, Erasmus MC, University Medical Center Rotterdam; Department of Health Sciences, VU University, Amsterdam

Reimbursement for attending a symposium?: No

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: No

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here:

Reviewer: 8

Comments:
This is a very important RCT in the field, given that no well conducted trial to date has yet been published that has tried to replicate the findings of the MAST trial (2013) by Albert H et al from Denmark. That trial's results were very controversial as it reported consistent and marked benefits across nearly all outcomes from antibiotics (compared to placebo) for chronic LBP patients with Modic type 1 changes on MRI, potentially representing a paradigm shift in the understanding of cause and treatment for this subgroup of patients - suggesting low grade bacterial infection in the spine). Since then several systematic reviews have been published suggesting there is evidence that low grade bacterial infection could contribute to low back pain. I am aware of one further small, non-registered but published trial from Iraq that showed positive results for antibiotics (Al-Falahi MA et al 2014) and two trials in progress, one in Australia (ACTRN12615000958583) and one in Iran (ACTRN12613001074785). With the current rising risk of antibiotic resistance caused by inappropriate use of antibiotics, and the prevalence of chronic low back pain with Modic changes seen on MRI, further high quality trial evidence is key to inform clinical practice.

This trial from Norway has many strengths, including its size and multi-centred nature, power to detect between group differences, extent of blinding, efforts to record treatment compliance and other medication use, the inclusion of patients with Modic type I (thus replicating the MAST trial) but also type II (to test whether it is Modic I that is the phenotype that identifies those who benefit from antibiotics). Other strengths are the ITT analyses, per protocol analysis and assessment of patient blinding. It was prospectively registered and as far as I can tell adheres well in terms of reporting the clinical outcome data listed in the registry and protocol paper. The processes of randomisation and allocation concealment are

explained clearly, with appropriate approaches to stratification by modic change type and previous disc herniation surgery.

Based on their results the authors make a strong conclusion that antibiotics (amoxicillin) did not provide a clinically important benefit and therefore the authors do not support the use of antibiotic treatment for chronic LBP and Modic changes.

I can see why they have drawn the above conclusion but I suspect the results may add to the international controversy rather than remove it. My reasons are:

1. The trial was powered to detect at least a difference of 4 points on the primary outcome of RMDQ (back pain-related disability, scored from 0 to 24 with 24 maximum disability) between antibiotic and placebo for each of the two modic change types. The SD was 5, hence the effect size that the trial was powered to detect was very large (0.8). They justify this based on the previous MAST trial findings AND a sense that in order to justify prescribing 100 days of antibiotics for these patients (with the risks associated) the effect would have to be large. Whilst I understand this, no other treatments are this effective for chronic LBP yet they are still recommended in guidelines, so this team set the bar very high. They found statistically significant differences between trial arms for Modic type I for RMDQ but no evidence of any effect on Modic type II but they do not discuss this. Is it possible that the addition of Modic type II patients in this trial served to 'dilute' the effect of antibiotic, if so, would a table in the appendix separately describing the results for those with Modic I compared to Modic II be useful in directing future research? Given the high power for the combined analysis (all modic change types together) they did find many statistically significant between group differences - in their primary outcome analysis of the RMDQ and in other outcomes (eg. function ODI, leg pain NRS, and health related quality of life EQ5D-5L). For RMDQ, patients started at about 13 and at 12 months those in the antibiotic group improved to about 9, versus about 11 in placebo. Whilst the improvement is nowhere near the dramatic benefit seen in the MAST trial (where patients started at 15 and improved to 5.7 by 12 months in the antibiotic group), the effect size observed between groups is about 0.32, so there is some evidence here that antibiotics might be similarly effective as other guideline recommended treatments (where effect sizes of 0.3 to 0.4 are typical). This effect size rises to 0.46 for those with Modic type I changes. In addition, when one compares the two trials in terms of NNT to achieve at least a 30% reduction in RMDQ (generally but not universally accepted as a minimally important change in RMDQ) the MAST trial's NNT was 4 (Modic I) and this trial's NNT is reported as 5(for combined Modic type I and II), but for those with only Modic type I I suspect the NNT would look better? I think the above points need discussed in this paper.

2. The authors themselves suggest that many patients allocated to the placebo group may have guessed they were in the placebo group and thus this may, to some extent, explain the results of the trial. One of the difficult-to-explain findings from the MAST trial was the unusual lack of ANY response, on average, in the placebo group, although the authors of that trial justified this on the basis that the patients all had longstanding pain, having tried and failed several other treatments and therefore were very unlikely to demonstrate a placebo response. I found that difficult to believe since the interaction alone in that trial (and the 1.5 hr education session all patients received) in a group of patients still seeking healthcare for their back pain would normally have been expected to effect some non-specific effects/placebo response. In this trial patients were recruited from those referred to participating hospitals (thus seeking healthcare) and those on the spinal surgery registry who had previous surgery but still had severe pain. There is no mention in the paper or appendices or CONSORT flow chart of the numbers identified and recruited from each of these methods of recruitment nor whether the participants in the trial from each recruitment method were similar - could this information be added? Also, what patient education did patients receive in this trial in addition to their tablets? How did it compare with the extensive 1.5 hr of education in the MAST trial? Finally, what previous treatments did patients have - did this differ, on average, between the groups?

3. Could the authors add a little to the discussion about the key similarities and differences in their trial versus the MAST trial? Examples of differences include the specific antibiotic used (in MAST they used co-amoxiclav), the somewhat more severely affected patient population in MAST (baseline RMDQ was 15 and many more had had previous spinal surgery - more than 50% in MAST versus only 21% in this trial), the inclusion of Modic type II and there may be others. Was the sample in this trial more heterogeneous than in MAST? Could this explain the results?

Minor points:
When was the SAP finalised, before or after the end of data collection? The SAP has a date of 2/11/2018 which seems after the end of the data collection - is that correct?

Could an appendix provide greater detail on the exclusions - the 310 that failed to meet the criteria, with 174 in the 'other' category.

Could the paper add explanation of where the other measures collected but not reported here, will be reported? The EQ5D is different between the groups hence the health economic analyses results will be interesting to see (since antibiotics are rather inexpensive, I suspect the quality of life difference will drive the economic results) - are they being reported separately? Also, where are the 12 month MRI scan results being reported?

Reference 16 - the protocol paper journal appears to be missing.

Additional Questions:
Please enter your name: Nadine Foster

Job Title: NIHR Professor of Musculoskeletal Health in Primary Care

Institution: Research Institute for Primary Care and Health Sciences

Reimbursement for attending a symposium?: Yes

A fee for speaking?: No

A fee for organising education?: No

Funds for research?: Yes

Funds for a member of staff?: No

Fees for consulting?: No

Have you in the past five years been employed by an organisation that may in any way gain or lose financially from the publication of this paper?: No

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: I collaborate with Prof Margreth Grotle from Oslo - one of the coauthors of this paper - on a different research topic. Margreth is a visiting professor at my UK University. I was not involved in any way in this trial.

Reviewer: 9

Comments:
GENERAL COMMENTS:

The authors present a randomized clinical trial evaluating the role of antibiotics in patients with chronic low back pain (CLBP) and Modic changes. This is an important study and the authors are to be commended for their performing it. I should also note that I am a healthy skeptic of the role of antibiotics as a treatment for patients with CLBP, including those with Modic changes. That said, I am concerned about the presentation of this study's results.

In the accompanying protocol documents submitted as supplementary material, the reason for including patients with Modic I and II changes in the current study makes sense. This isn't reflected in the paper and caused me considerable confusion. Prior work has focused on those with Modic I changes and while expanding to include those with Modic II changes is reasonable, I was confused and then bothered by a lack of presentation of the subgroup with Modic I changes. Indeed, even in the supplementary tables, there is no presentation of primary and secondary outcomes among those with Modic I changes. I guess the authors will argue that these weren't specified in their analysis plan as key outcomes, but I think that is a problem.

I also found the emphasis on a difference of 4 or greater in the RMDQ to be problematic. In reviewing the authors' analysis plan, I couldn't find mention of this as part of the primary outcomes assessment. Rather I found mention of a 30% or greater improvement in RMDQ in the analysis plan. The mention of the DMDQ difference of 4 is included in the power calculations section. Thus, the very large emphasis on discounting the results favoring the intervention based upon the primary outcome not meeting this difference seems overplayed.

In trying to digest these results, I came to the following conclusions. First, it appears that those patients receiving antibiotics had a statistically significant improvement in the primary outcome and several secondary outcomes compared to those receiving placebo. Moreover, the magnitude of the benefit appeared limited to those with Modic I changes (as mentioned this is based upon only the primary outcome, not any secondary ones since they aren't reported). Second, the improvement seen in this study is considerably less than the prior RCT. Third, major side effects of the antibiotics were pretty limited and if the magnitude of the benefit was sufficiently large, many providers (including myself) would use them. Finally, it isn't clear how this treatment compares to other commonly used therapies for CLBP, but the magnitude of the effect of antibiotics seems to be comparable to other treatments in use.

Putting all this together, I find it harder to entirely discount the small but favorable results seen in this study than the authors have. Before further consideration of these results for publication in BMJ, I would like to see all the primary and secondary outcomes in those with Modic I changes. I think that a more nuanced interpretation of the results would likely result. I suspect this won't fully satisfy those who favor this treatment as well as those who don't.

SPECIFIC COMMENTS:

1. Abstract, Page 3, Lines 82-83: I presume that most BMJ readers won't know the potential significance of Modic changes and the theory behind antibiotic treatment.

2. Abstract, Page 3, Lines 96-97: In reading through the protocol attachment, I had difficulty identifying a statement where the MCID between group difference of 4 was part of the primary outcome assessment. Rather in section 3.4.1 it says, "A clinically important change will be defined as a 30% reduction of the individual's baseline score." I could not find reference to the current sentence in the abstract in sections describing primary outcomes.

3. Abstract, Page 3, Lines 99-104: As noted above, the authors do not report the percent of patients with a 30% reduction in RMDQ. From Table S4, it appears that there is a statistically significant difference between the two study groups (48% vs. 29%, p=0.01). Seems like this should be included in the paper and abstract.

4. Abstract, Page 3, Lines 100-102: It isn't clear to me why the difference in outcomes among Modic I and II groups isn't highlighted in the conclusions. Doesn't this support the hypothesis under investigation? I'd argue that while a secondary outcome based upon it being a subgroup, in a differently designed study it would be the key result. At a minimum, the conclusion is that this requires further study in an adequately designed trial of individuals with Modic I changes.

5. Abstract, Page 3, Lines 106-107: It appears that the authors are downplaying the results of their primary outcomes. Though the differences may not be large, the intervention group had superior outcomes on the primary measure, the clinically important change outcome, and several key secondary measures. Why isn't this good enough?

6. Introduction, Page 3, Lines 113-115: It would be helpful to comment upon how common each of these findings are and how reliably they can be identified.

7. Introduction, Page 3, Lines 129-130: As per the prior comment, it would be helpful to comment upon the breakdown of the different Modic categories.

8. Introduction, Page 3, Lines 132-134: Given that the one study cited found benefit in individuals with type I Modic changes, why include both type I and II?

9. Methods, Page 4, Lines 183-192: Was data collected on use of treatments for back pain during the 1 year study period? Similarly, was data collected on use of antibiotics for other indications during the study period?

10. Methods, Page 4, Lines 209-212: As previously noted, I was unable to find mention of this in the protocol supplements. The MCID for the RMDQ is often cited as greater than 2. What was the basis for a between group difference of 4?

11. Results, Page 6, Line 257-259: I think it is worth mentioning that almost all secondary outcomes in Table 2 also showed a statistically significant difference favoring the intervention group. I also think the authors should state in the results whether the as-treated outcomes were similar to the ITT ones.

12. Results, Page 6, Lines 268-270: I'm not an expert on blinding analyses, but it appears that the control group was able to know their treatment assignment more than for the intervention group. This bias doesn't appear to be stated as such in the limitations paragraph of the discussion.

13. Discussion, Page 6, 273-275: Unfortunately, I find this conclusion to lack the necessary nuance for a BMJ publication. The study showed statistically significant changes favoring the intervention on primary and a range of secondary outcomes. Moreover, the magnitude of the benefit seen appears to have been limited to those with Modic I changes, as expected. One can then question the whether the the difference in outcomes is of sufficient magnitude, but I am concerned that this statement does not adequately reflect the study analyses and findings.

14. Discussion, Page 6, 276-279: Similar to my prior comments, one could take a very different perspective on the observed secondary outcomes. Additionally, there is no mention of as-treated results or the greater impact in those with Modic I changes.

15. Discussion, Page 6, 281-283: Did the other trial include the same percentage of patients with Modic I and II changes. Given that the current study shows all of the benefit appears to be in those with Modic I changes, the combined outcomes may underestimate the true benefit of the intervention. It appears that there was a difference of 2.3 in the current study among those with Modic I changes. This is less than 8.3 in the other trial, but the direction is the same. Then one may approach somewhat differently why the current study was associated with less benefit. Differences in the antibiotic regimen are highlighted, but others should also be explored.

16. Discussion, Page 6, 295-298: Burying this in the limitations paragraph seems a bit disingenuous to me. At a minimum, these results support further study in those with Modic I changes as the primary outcome.

17. Discussion, Page 6, 298-300: If the study was adequately powered for the subgroup of Modic I changes, why don't the authors present this subgroup analysis in an additional table in the paper?

18. Discussion, Page 7, 310-311: In what way do these decrease the generalizability of the study findings? Does the population studied not reflect those with Modic I or II changes in other studies?

19. Discussion, Page 7, 320-321: These outcomes are not presented in the results section. Please add them or remove mention here.

20. Discussion, Page 7, 328-330: Revise per prior comments.

21. Page 3, Lines 10-11: I've provided some suggestions for revising this paragraph. I'd suggest that this second sentence isn't particularly helpful and that it can be deleted.

22. Table S4: It appears that these are the results that pertain to the comment above about not being mentioned in the results section.

23. Table S7: I think mention of the per-protocol findings in the results section as a line of text would be helpful.

Additional Questions:
Please enter your name: Steven Atlas

Job Title: Director, Practice-Based Research & Quality Improvement

Institution: Massachusetts General Hospital

Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this paper?: No

If you have any competing interests <A HREF='http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/declaration-competing-interests'target='_new'> (please see BMJ policy) please declare them here: I am the section editor for UpToDate for back and neck content and receive royalty payments for this work.