Statistical analysis plan for analysis of MRI-defined effect-modifiers in the AIM trial

Statistical analysis plan for:

MRI-defined effect-modifiers of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM trial)

Publication date: September 23, 2019

Document version: 1.0

EudraCT Number: 2013-004505-14

This document is a supplement to the AIM trial protocol [1] and comprises a statistical analysis plan (SAP) for an article with tentative title "MRI-defined effect-modifiers of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM trial)". Separate SAPs exist for other articles based on the AIM trial [2]. The current SAP was prepared based on guidelines for Statistical analysis plans in clinical trials [3], and after results for treatment efficacy in the main trial and in clinical subgroups were available.

Scientific board

John-Anker Zwart, professor dr.med
Kjersti Storheim, professor, PT, PhD
Jens Ivar Brox, professor dr.med
Ansgar Espeland, professor dr.med
Margreth Grotle, professor, PT, PhD
Olav Lutro, MD
Anne Froholdt, MD, PhD
Lars Grøvle /Anne Julsrud Haugen, MD, PhD
Audny Anke, professor dr.med
Øystein P Nygaard, professor dr.med
Jan Sture Skouen, professor dr.med
Christian Hellum, MD, PhD
Benedicte A Lie, PhD
Karianne W Gammelsrud, MD, PhD
Thor Einar Holmgaard (patient representative)

Statistical advisor

Jörg Aßmus, PhD

Sponsor

Oslo University Hospital

Coordinating investigator

John-Anker Zwart, professor dr.med
FORMI, OUS Ullevål, Bygg 37B, Postboks 4956 Nydalen, 0424 Oslo
Tel: 22 11 86 21
E-mail: j.a.swarz@medisin.uio.no

Contributors to current SAP

Ansgar Espeland, professor dr.med
Per Martin Kristoffersen, MD, PhD-student
Lars Christian Bråten, MD, PhD-student
Lars Grøvle, MD, PhD
Nils Vetti, MD, PhD
Christian Hellum, MD, PhD
Jörg Aßmus, PhD, statistician

Signature page

Title **MRI-defined effect-modifiers of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM trial)**

Trial registration: ClinicalTrials.gov, ID: NCT02323412

EudraCT no: 2013-004505-14

I hereby declare that I have reviewed and approved the statistical analysis plan:

To be signed by person writing the SAP, senior statistician responsible, project manager, and coordinating investigator.

Name	Title	Role	Signature	Date
Ansgar Espeland	Professor, dr.med	Main author of current SAP		25/09-19
Per M Kristoffersen	MD, PhD-student	Contributor to current SAP		25/09-19
Jörg Aßmus	PhD	Senior statistician responsible		25/09-19
Kjersti Storheim	Professor, PT, PhD	Project manager		3/10-19
John-Anker Zwart	Professor dr.med	Coordinating investigator		30/9-19

Table of contents

Abbreviations and definitions

Abbreviation or term	Explanation
AIM	Antibiotics in Modic changes
CI	Confidence interval
CSF	Cerebrospinal fluid
ITT	Intention to treat
MCs	Modic changes
MRI	Magnetic resonance imaging
NRS	Numerical rating scale
ODI	Oswestry disability Index
RMDQ	Roland Morris disability questionnaire
SAP	Statistical analysis plan
STIR	Short tau inversion recovery

Background and rationale

Modic changes (MCs) are MRI signal changes in the vertebral bone marrow extending from the endplate, and are classified as type I (oedema type), II (fatty type) and III (sclerotic type, less common) based on standard T1- and T2 weighted sequences [4].

The association between MCs and low back pain is inconsistent [5], but it seems that particularly type I MCs might be related to pain [6-8]. Different explanations for MCs are hypothesized [9].

One hypothesis is that low back pain and MCs are caused by an infection with Cutibacterium acnes in the intervertebral disc. Based on this hypothesis, the AIM trial investigated the effect of Amoxicillin in patients with chronic low back pain and type I or II MCs. The trial is described in more detail below.

The present study of MRI-defined effect-modifiers is part of the AIM trial. The rationale for this study is the hypothesis that antibiotic treatment is more effective when the MC displays more extensive oedema signals or is larger, potentially reflecting a more biologically active lesion. The study is based on standard T1/T2 series and STIR. STIR is sensitive to oedema signals, both in MCs classified as type I and MCs classified as type II, based on T1/T2 series. More marked oedema will appear on T1/T2, and defines the MC as oedema type I. The AIM trial found a better, but not clinically important, treatment effect in the type I versus the type II MC group [10]. This finding further motivated the present study performed to assess whether baseline MRI results (beyond MC type group) might suggest improved, clinically relevant effect of antibiotic treatment at one-year follow-up.

The AIM trial

The AIM trial was a six centre, randomised, parallel-group, placebo-controlled trial on the effects of three months of treatment with Amoxicillin in chronic low back pain patients with type I or II MCs at the level of a prior lumbar disc herniation (verified on MRI during the preceding 2 years) [1]. Randomisation to either amoxicillin or placebo was stratified by MC type group (I/II) and prior disc surgery (yes/no) with a 1:1:1:1 allocation and random block sizes of 4 and 6.

Patients were classified in the MC type I group if they had primary and / or secondary type I MCs at a level with previous disc herniation. In primary type I MCs, type I is the most extensive type at the

evaluated endplate. In secondary type I MCs, type I is present, but another type is more extensive. Patients were classified in the MC type II group if they had type II MCs, but not any type I MCs (primary or secondary), at a previously herniated disc level.

MCs with maximum height < 10% of vertebral body height or diameter ≤ 5 millimetres did not qualify for inclusion in the trial. These size criteria concern total MC size, not the size of individual MC types.

The primary outcome in the AIM trial was pain-related disability measured by the RMDQ (score range 0 to 24) at one-year follow-up. Secondary outcomes included low back pain intensity on a 0-10 NRS and pain-related disability assessed by the ODI [1]. The minimal clinically important between-group difference in mean RMDQ score was defined as a difference of 4.

The AIM trial was designed to evaluate the treatment effect in the total sample as well as separately in each MC type group (I/II). In each MC type group, it was designed to detect ($\beta = 0.1$, two-sided $\alpha = 0.05$) a difference of 4 (SD 5) in mean RMDQ score between the amoxicillin group and the placebo group at one-year follow-up. This required 66 patients in each MC type group. In total 180 patients were actually included, 118 in the MC type I group and 62 in the MC type II group; 155 completed the trial without major protocol deviations [10]. See the trial protocol for further details regarding sample size calculation and trial methods [1].

The AIM trial showed a statistically significant – but not clinically important – effect of Amoxicillin in the type I MC group, but not in the in the type II MC group [10]. A clinical subgroup study (not published yet) indicated better effect in younger patients (aged < 40 years).

Objective of the present study

The objective of the present study is to evaluate potential MRI-defined effect-modifiers of amoxicillin treatment in patients with chronic low back pain and type I or II MCs at the level of a previous lumbar disc herniation.

Hypotheses for the present study

Both main hypotheses for the present study were pre-specified in the AIM-trial protocol [1].

1. Hypothesis 1 (prioritized as the sixth hypothesis F in the AIM trial protocol) is:

   In the antibiotic treatment group, high signal from MCs on STIR at baseline MRI predicts lower RMDQ-score at one-year follow-up.

2. Hypothesis 2 (not listed among the eight prioritized hypotheses in the AIM-trial protocol) is:

   A positive effect of antibiotic treatment at one-year follow-up is more likely when the baseline MCs contain more type I than type II or are larger.

Hypothesis 1 is prioritised before hypothesis 2, since this order conforms to the AIM-trial protocol. Hypothesis 2 is part of a broader hypothesis in the AIM protocol, that a positive effect of antibiotic treatment at one-year follow-up is more likely when baseline MCs a) contain more type I than type II based on conventional T1/T2 weighted MRI and STIR, b) are more intense on STIR, or c) are larger.

Statistical principles

The present SAP was developed after locking the AIM database, including the relevant MRI findings, and after analysing clinical outcomes and clinical subgroup effects. All analyses described in this SAP are considered *a priori* analyses in that they have been defined in this SAP and / or the AIM trial protocol. Any *post hoc* analyses will be identified as such in the article. All statistical tests will be 2-sided and the nominal p value will be reported. All CIs presented will be 95% and 2-sided. The assumption of normal distribution will be checked by visual inspection of a QQ-plot.

The general statistical significance level is set to 0.05, as advised for 'hypothesis-setting studies' to avoid type II errors [11]. Yet, when *testing* hypothesis 1 (the sixth hypothesis F in the AIM protocol) we will use a Bonferroni corrected significance level of 0.05/6 = 0.008. We will use further pre-specified criteria (see below) to assess the credibility and clinical relevance of our results. Note, that the Bonferroni adjustment may be too conservative, since the tests are not independent. This strengthens positive findings.

Analysis populations

We will perform all analyses of potential MRI-defined effect-modifiers both in the ITT population (n=180) and in the per-protocol population (n=155). The per-protocol analysis will be the primary analysis. The ITT analysis will be a supportive analysis. The ITT population is all patients randomized to amoxicillin or placebo. The per-protocol population is all patients completing the study without major protocol deviations.

Major protocol deviations were defined as (a) intake of < 80% of the pills (amoxicillin or placebo), (b) pause of the study medication for ≥ 2 weeks (in the antibiotic group: without other 'relevant' antibiotic treatment in that period), (c) 'relevant' antibiotic treatment in the placebo group for ≥ 4 continuous weeks between baseline and one-year follow-up, and (d) back surgery during the one-year follow-up. 'Relevant' antibiotic treatment was antibiotic treatment likely to affect a C. acne discitis. Further events registered as major protocol deviations were faulty inclusion (2 patients treated with antibiotics last month prior to inclusion), both amoxicillin and placebo given to patient by mistake (1 patient), and spondyloarthritis diagnosed during follow-up (1 patient).

Trial population

The main report of the AIM trial [10] includes a flow diagram with number of patients assessed for eligibility, number of patients included and randomised, number of screened patients not included, and the reason for non-inclusion. The diagram also shows the number of ineligible patients who were randomised (n=2), with reasons for ineligibility. It further shows lost to follow-up, withdrawal from follow-up and treatment non-completion for each treatment arm with timing and respective reasons. This study's report will refer to the prior flow diagram and - as a minimum – include the number of patients randomised and the number in the per-protocol population for each treatment arm.

MRI variables

This study is based on 1.5 T baseline MRI data from conventional T1- and T2-weighted fast spin echo series and STIR sequences. Potential MRI-defined effect-modifiers are defined based on MRI findings at the index level(s) only, i.e. the level(s) with type I and/or type II MCs and a prior disc herniation.

We have predefined one composite STIR variable and one composite T1/T2 variable that we will use to test hypotheses 1 and 2, respectively (Table 1). Each composite variable was established by 'meaningful grouping' (a non-statistical approach) [12] of underlying variables based on earlier research and clinical plausibility (Table 1). This was done prior to analysing the effect of any MRI variable on outcome, but not blinded to the variables' distribution in our sample. To reduce the problem of multiple testing, we did not first perform multiple explorative analyses of the underlying variables to identify optimal composite variables. Such analyses may be considered *post hoc*.

We have predefined four underlying STIR variables and three underlying T1/T2 variables. The STIR variables are volume (categorised, not measured), maximum height (measured), maximum intensity (measured) and presence both superior and inferior to disc (yes/no) of visible MC related high STIR signal compared to normal vertebral bone marrow (Table 1). The T1/T2 variables are type I degree (categorised), volume (categorised) and maximum height (measured) of MCs on T1/T2 (Table 1).

Each of the seven underlying variables also defines separate subgroups that will be examined in further explorative hypothesis-setting analyses (Table 1).

All conclusive MRI findings are based on two or three radiologists' independent evaluations.

Description of baseline characteristics

In the main report of the AIM trial [10], patients in each treatment group are described with respect to age, gender, body mass index, smoking, education level, comorbidity, emotional distress, fear-avoidance beliefs, leg pain, duration of back pain, former disc herniation surgery, physical work load, compensated work injury or sick leave, concomitant medication use, and level(s) with both MC and previously herniated disc.

The report of this study will refer to the prior description and - as a minimum - include age, gender and all MRI variables for each treatment group (Table 2). We will summarise categorical variables by numbers and percentages, and continuous variables by mean and SD (if normal distribution) or median and interquartile range (if skewed distribution). We will note the clinical importance of any imbalance in baseline data between the treatment groups without statistical testing.

Analyses

Each potential MRI-defined effect-modifier (Table 1) will be analysed by its interaction term with treatment. ANCOVA will be performed with RMDQ score at one year as dependent variable adjusted for baseline RMDQ score, and adding the potential effect modifier, treatment group and the interaction term (effect modifier*treatment group). Missing values of RMDQ will be substituted with imputed values from the multiple imputation performed in the primary AIM trial analysis [2].

In all analyses, we will adjust for age and for former disc herniation surgery (yes/no) (a stratification variable in the randomization), but not for 'MC type group' (the other stratification variable), since this variable overlaps considerably with the potential MRI-related effect-modifiers we want to assess.

We will use the baseline value for all potential effect modifiers and covariates. Within each category (subgroup) of categorical variables, we will report the treatment effect predicted in the interaction term analysis (and not perform a separate stratified analysis). The minimal clinically important difference in mean RMDQ score between the treatment groups remains defined as 4.

We will present results in a forest plot of effect estimates for each potential effect modifier with 95 % CIs and the p value for the interaction (effect modifier*treatment group). The overall treatment effect will be marked on the plot [13]. The plot will also show the treatment effect within subgroups.

If one or both composite MRI variables modify the treatment effect with $p < 0.05$ in the interaction test, we will add both of them and their interaction terms with treatment in the same ANCOVA analysis [14]. In this analysis, we will also adjust for age and former disc herniation surgery and add age*treatment group, since age modified the treatment effect in our prior analyses. We may also consider an analysis with all MRI variables and interaction terms added, as would be ideal to assess the independency of any effect modification, but our study is poorly powered for such an analysis.

We will repeat the analyses with LBP intensity and ODI score at one year as dependent variables to assess if any subgroup effect is consistent across related outcomes, as recommended [14].

Criteria for evaluating the results of the present effect modifier analyses

When interpreting the plausibility of any effect modifiers, we will use the following criteria for evaluating subgroup analyses [14] (currently known answers for present analysis in parentheses):

Design

Is the subgroup variable a characteristic measured at baseline or after randomisation?* (At baseline)
Is the effect suggested by comparisons within rather than between studies?
Was the hypothesis specified a priori? (Yes)
Was the direction of the subgroup effect specified a priori?* (Yes)
Was the subgroup effect one of a small number of hypothesised effects tested?

Analysis

Does the interaction test suggest a low likelihood that chance explains the apparent subgroup effect?
Is the significant subgroup effect independent?*

Context

Is the size of the subgroup effect large?
Is the interaction consistent across studies?
Is the interaction consistent across closely related outcomes within the study?*
Is there indirect evidence that supports the hypothesised interaction (biological rationale)?

*New criteria.

Table 1 – Predefined potential MRI-defined effect-modifiers and analyses

Hypotheses and rationale	Predefined variables and subgroups
STIR signal extent and intensity	
Hypothesis 1: In the antibiotic treatment group, high signal from MCs on STIR at baseline MRI predicts lower RMDQ score at one-year follow-up (hypothesis F in AIM trial protocol [1]). Rationale: Oedema type MCs may be related to pain [6-8]. Bone marrow oedema can also indicate a biologically "active" lesion potentially responsive to treatment. STIR is sensitive to oedema, visible as high signal intensity. We therefore expect better effect of antibiotic treatment (i.e. lower RMDQ score at one year) in subgroups with a larger volume, maximum height, and maximum intensity of the region(s) with MC related high signal on STIR. We regard volume and maximum height to be different, yet related, characteristics of the lesions. Based on our clinical experience that oedema both superior and inferior to the disc is often found in biopsy-verified bacterial spondylodiscitis, we expect that such a finding might make bacterial discitis and a treatment effect more likely. *The categorisation of STIR intensity helps to avoid exclusion of patients without a measured value. The <25 % category includes MCs with no conclusive STIR signal increase (or decreased STIR signal) and thus no measured intensity values. These values were likely to have been <25 % had they been measured, since the value was <20 % for >90 % of intensity measurements reported by only one observer (as the other observers found no visual signal increase and therefore did not measure the intensity). **We may also analyse STIR intensity as a continuous variable (with values imputed for those missing) if the categorised variable is found to modify the treatment effect.	We define MC related high STIR signal as visible high signal compared to normal vertebral bone marrow, formed and located as an MC and / or located in or abutting a region with MC on T1/T2. **(i) STIR - a composite STIR variable** we will use to test hypothesis 1 - is based on these variables: (ii) **STIR volume** – largest volume of high STIR signal in % of vertebral body marrow volume, visually estimated (not measured) and scored as 0, 1 (<10 %), 2 (<25 %), 3 (25-50 %) or 4 (>50 %). We will analyse scores 0-1 versus 2 versus 3-4. (iii) **STIR height** - maximum height of region with high STIR signal, measured and re-calculated as a proportion of vertebral body marrow height. We will analyse this proportion both as a continuous variable and dichotomized (≤0.50 versus >0.50). (iv) **STIR intensity** - maximum intensity of the high STIR signal, measured and re-calculated into a percentage on a scale from normal vertebral body marrow intensity (0 %) to CSF intensity (100 %), and categorised in the analyses (<25 % versus 25-40 % versus >40 %). See comments left.*, ** (v) **STIR sup/inf** – presence of high STIR signal both superior and inferior to disc (yes/no). **(i) STIR** is categorized as **STIR 3**(volume ≥25 % AND height >0.50 AND intensity ≥25 % AND yes for sup/inf), **STIR 2** (not STIR 3 or STIR 1) and **STIR 1** (volume <25 % AND intensity <25 %). Variables (ii-v) will be used separately in further explorative analyses, as described above.
MC type and extent on T1/T2	
Hypothesis 2: A positive effect of antibiotic treatment at one-year follow up is more likely when baseline MCs contain more type I than type II or are larger. Rationale: As noted above, oedema type I MCs	**(a) MC - a composite T1/T2 variable** we will use to test hypothesis 2 - is based on these variables: (b) **MC type I degree** -categorised as type I major (primary type I both superior and inferior to disc), type I minor (primary or secondary type I, but not

may be related to pain [6-8] and oedema may indicate a biologically "active" lesion potentially responsive to treatment. We therefore expect that MCs containing type I as the primary (most extensive) type would predict a better effect of antibiotic treatment (i.e. lower RMDQ score at 1 year). Based on clinical experience, we expect that signs of oedema both superior and inferior to the disc might make a bacterial discitis and an effect of antibiotic treatment more likely (see above). We also expect large lesions to be more clinically "active" and responsive to treatment than lesions with limited extent. Prior to the development of this SAP, the AIM trial had shown better (but not clinically important) effect of antibiotic treatment in the MC type I group than in the MC type II group. We expect further improved effect in patients with a more pronounced type I affection. * The type II only category is the same for variable (a) and variable (b).	type I major) and type II only (not type I). (c) **MC volume** – largest volume of MC (including all MC types at given endplate) in % of vertebral body marrow volume, visually estimated and scored as 0, 1 (<10 %), 2 (<25 %), 3 (25-50 %) or 4 (>50 %). We will analyse score 1 versus 2 versus 3-4 (largest index level MC volume cannot be 0). (d) **MC height** - maximum MC height (including all MC types), measured and re-calculated as a proportion of vertebral body marrow height. We will analyse this proportion both as a continuous variable and dichotomized (≤0.50 versus >0.50). **(a) MC** is categorized into **type I++** (type I major AND volume ≥10 % AND height ≥25% - OR - primary type I AND volume ≥25 % AND height >0.50), **type I+** (primary or secondary type I but not type I++) and **type II only,** see comment* left. Variables (b-d) will be used separately in further explorative analyses, as described above.

Table 2 – Distribution of baseline MRI variables by treatment group

Variable	Amoxicillin group (N = 89)		Placebo group (N = 91)	
	n	%	n	%
(i) STIR – composite variable				
STIR 3				
STIR 2				
STIR 1				
(ii) STIR volume – maximum score				
0				
1				
2				
3				
4				
(iii) STIR height – proportion of vertebral body marrow height				
Mean, SD or median, interquartile range				
<0.25				
0.25-0.50				
>0.50				
(iv) STIR intensity - % increase from normal vertebral body marrow intensity (0%) to CSF intensity (100%)				
Mean, SD or median, interquartile range				
<25				
25-40				
>40				
(v) STIR sup/inf – STIR signal increase both superior and inferior to disc				
Yes				
No				
(a) MC - composite variable				
Type I++				
Type I+				
Type II only				
(b) MC type I degree - categories				
Type I major				
Type I minor				
Type II only				
(c) MC volume – maximum score (cannot be 0 at index level)				
1				
2				
3				
4				
(d) MC height – proportion of vertebral body marrow height				
Mean, SD or median, interquartile range				
< 0.25				
0.25-0.50				
> 0.50				
Index level(s) with MC and previous disc herniation				
L2/L3				
L3/L4				
L4/L5				
L5/S1				

References

1 Storheim K, Espeland A, Grovle L et al (2017) Antibiotic treatment In patients with chronic low back pain and Modic changes (the AIM study): study protocol for a randomised controlled trial. Trials 18:596

2 Bråten LC RM, Espeland A, Storheim K, Zwart JA, Hellum C, Aßmus J (2018) Statistical analysis plan (SAP) for clinical outcomes in the AIM-study: A randomized trial of antibiotic treatment in patients with chronic low back pain and Modic Changes (the AIM-study), ClinicalTrials.gov

3 Gamble C, Krishan A, Stocken D et al (2017) Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. JAMA 318:2337-2343

4 Modic MT, Steinberg PM, Ross JS, Masaryk TJ, Carter JR (1988) Degenerative disk disease: assessment of changes in vertebral body marrow with MR imaging. Radiology 166:193-199

5 Herlin C, Kjaer P, Espeland A et al (2018) Modic changes-Their associations with low back pain and activity limitation: A systematic literature review and meta-analysis. PLoS One 13:e0200677

6 Hanimoglu H, Cevik S, Yilmaz H et al (2019) Effects of Modic Type 1 Changes in the Vertebrae on Low Back Pain. World Neurosurg 121:e426-e432

7 Maatta JH, Karppinen J, Paananen M et al (2016) Refined Phenotyping of Modic Changes: Imaging Biomarkers of Prolonged Severe Low Back Pain and Disability. Medicine (Baltimore) 95:e3495

8 Splendiani A, Bruno F, Marsecano C et al (2019) Modic I changes size increase from supine to standing MRI correlates with increase in pain intensity in standing position: uncovering the "biomechanical stress" and "active discopathy" theories in low back pain. Eur Spine J 28:983-992

9 Albert HB, Kjaer P, Jensen TS, Sorensen JS, Bendix T, Manniche C (2008) Modic changes, possible causes and relation to low back pain. Med Hypotheses 70:361-368

10 Bråten LCH (2019) 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. The British Medical Journal In print

11 Kent P, Keating JL, Leboeuf-Yde C (2010) Research methods for subgrouping low back pain. BMC Med Res Methodol 10:62

12 Song MK, Lin FC, Ward SE, Fine JP (2013) Composite variables: when and how. Nurs Res 62:45-49

13 Cuzick J (2005) Forest plots and the interpretation of subgroups. Lancet 365:1308

14 Sun X, Briel M, Walter SD, Guyatt GH (2010) Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. BMJ 340:c117