Supplementary Appendix

Table of Contents

Methods

Statistical Analysis

In all analyses and descriptive statistics of the primary endpoints (RMDQ consisting of 24 items), we performed a mean-imputation in cases with less than 30% of the items missing at any time point. We then performed a multiple imputation in those cases where more than 30% of the questions were missing at any time point for the analyses of treatment comparison. The imputation model included a broad group of variables associated with the primary outcome,[1] as described in the Statistical analysis plan. We defined drug effect as the difference between the mean effect of amoxicillin and the mean effect of placebo.

In the primary analysis we compared mean RMDQ scores at one year between the two treatment groups in the whole intention-to-treat population using analysis of covariance (ANCOVA), adjusted for baseline RMDQ score and the stratification variables (Modic Changes type and previous disc surgery).

We also performed analyses of secondary outcome measures (Oswestry Disability Inex, low back pain intensity and health related quality of life) in a similar manner as the primary outcome, using multiple imputation models with significance level adjusted to 0. 0167 as described in the Statistical Analysis Plan.[2]

We performed a per-protocol analysis for the primary outcome, where all patients with a major protocol deviation were excluded. Major protocol deviation is defined as noncompliance (taking less than 80% of the prescribed pills), incorrect enrollment, back surgery during the trial period, pause of the study medication (both treatment groups), or intake of antibiotics (in the placebo group) as reported in table 5 in the Statistical Analysis Plan.[2] However, patients who stopped the study medication due to clinical reasons (i.e. adverse events not related to protocol deviation) were included in the per-protocol population.[3] Compliance was defined as the percent of pills that the patient has taken out of the planned number of pills. It was calculated based on data on counts of all the pills that the patients returned at end of treatment (table S1).

Minor protocol deviation is defined as compliance less than 95% but more than 80%, any treatment for back pain (initiated at any time point between baseline and one year follow up), pause of study medication (both treatment groups) or intake of antibiotics (in the placebo group) as reported in table 5 in the Statistical Analysis Plan.[2] In cases where data for compliance was missing, we relied on the weekly patient reported compliance questionnaire (table S1). Out of 28 patients with missing information on counts of returned pills, 23 patients had reported taking tablets less than 80% of the days (or had missing information about these reported numbers) and were defined as noncompliance, and hence also as having a major protocol deviation. Treatment non-completion (reported in figure 1 in the main manuscript) is defined as either stopping study medication due to adverse events or non-compliers (the distinction between these two is relevant for deciding if the patient were included in the per-protocol population or not, as described above).

We calculated, separately for each treatment group, the number and percentage of patients with 1) one or more adverse events, 2) one or more grade 2 adverse events, 3) serious adverse events or grade 3–4 toxicity and 4) symptoms of diarrhea, abdominal pain, rash, candida infection. For each of these four categories of events/symptoms, we also report the number of drug related adverse events (defined as possible/probable/definite relationship to treatment) and adverse events with unrelated/unlikely relationship to treatment. For all these calculations we present events/symptoms descriptively without any formal statistical testing.

Further analyses included responder analyses (improved >30, >50 and > 75% compared to baseline value, excluding patients with >30% items missing of the RMDQ) and Linear mixed-effects (LME) models as described in the Statistical Analysis Plan.[2]

Unweighted kappa was calculated for agreement between the study radiologists on trial eligibility based on magnetic resonance image findings of disc herniation and Modic Changes, using the categories eligible for the Modic Changes type I group, eligible for the Modic Changes type II group, or not eligible.

The graphics were derived using Matlab 9.0 (Natick, MA).

Figure S1 - List of inclusion and exclusion criteria

INCLUSION CRITERIA

- Age between 18 and 65 years
- LBP of > 6 months duration in the area below the 12th rib and above the gluteal folds with a Numerical Rating Scale (NRS) pain intensity score of ≥ 5 (mean of three 0–10 NRSs; current LBP, the worst LBP within the last 2 weeks, and usual/mean LBP within the last 2 weeks).
- MRI-confirmed lumbar disc herniation within the preceding 2 years.
- Type I and/or type II MCs in the vertebral body marrow at the same level as the previously herniated disc. For patients with previous surgery for disc herniation, the MCs has to be located at an operated level.
- Written informed consent

EXCLUSION CRITERIA

- Allergy to penicillin or cefalosporins
- Allergy/hypersensitivity to any of the excipients of the study drug
- Current pregnancy or lactation
- Kidney (creatinine) or hepatic (ALT/AST) laboratory values above the normal range
- Phenylketonuria (Følling's disease)
- Mononucleosis or leukaemia
- Any specific diagnosis that may explain patient's low back symptoms (e.g. tumor, fracture, spondyloarthritis, infection, spinal stenosis).
- Previous low back surgery (L1 – S1) for reasons other than disc herniation (e.g fusion, decompression, disc prosthesis).
- Surgery for disc herniation within the last 12 months
- Previous surgery for disc herniation, but MCs located at level(s) that has/have not been operated on only.
- Reservation about the intake of gelatine (the capsules contains gelatine, which among other things is produced by ingredients derived from pigs)
- Regular use of glucocorticoids
- Regular use of opioids with the exception of codeine and tramadol
- Not understanding Norwegian language
- Unlikely to adhere to treatment and/ or complete follow-up (e.g serious ongoing psychiatric disease, drug abuse, plans to move)
- Antibiotic treatment within the preceding one month before treatment start
- Contraindications to MRI (e.g. cardiac pacemaker electrodes, metal implant in eye or brain, claustrophobia).
- Unwilling to participate

LBP=low back pain; MRI=magnetic resonance image; MCs=Modic changes; ALT=alanine transaminase; AST=aspartate transaminase

79 **Table S1 - List of trial measurements with timing**

Trial measurements	Timeline
Primary outcome	
– **Roland-Morris Disability Questionnaire (RMDQ), Norwegian validated version, range 0–24, with higher score indicating more severe pain and disability, measured at one year follow-up (Primary endpoint).** [4,5]	Reported at 0, 3, 6, 9 and 12 months
Secondary outcomes	
– **Oswestry Disability Index (ODI) 2.0, range 0–100.** [6] **(Key supportive endpoint)**	Reported at 0, 3 and 12 months
– **Low back pain intensity (mean of three Numeric Rating Scales (NRSs, range 0-10); current LBP, the worst LBP within the last 2 weeks, and usual/mean LBP within the last 2 weeks (for weekly reports during the intervention period; the wording "last 2 weeks" was replaced by "the last week")**[7] **(Key supportive endpoint)**	Reported every week during treatment period and at 6, 9 and 12 months
– **Health-related quality of life (EuroQoL-5D, version 2.0)**[8] **(Key supportive endpoint)**	Reported at 0, 3 and 12 months
Further outcomes	
– **Leg pain intensity (NRSs, range 0–10) last week** [6]	Reported at 0, 3 and 12 months
– **Hours with LBP during the last 4 weeks**	Reported at 0, 3 and 12 months
– **Patients' satisfaction (5-point Likert scale)**	Reported at 3 and 12 months
– **Global perceived effect (7-point Likert scale)**	Reported at 3 and 12 months
– **Work status including days with sick leave**	Reported at 0, 3 and 12 months by care providers and every month during the trial period by the patient
– **Co-interventions (other pharmacological (ATC-coded) and non-pharmacological treatments)**	Reported at 0, 1,2, 3 and 12 months by care providers and every month during the follow up phase by the patient
Other trial measurements	
– **Emotional distress (Hopkins Symptom Checklist–25)** [9]	Reported at baseline
– **Fear-avoidance beliefs Questionnaire (FABQ)** [10]	Reported at baseline
– **Subjective health complaints (SHC)** [11]	Reported at baseline
– **Background information (age, gender, BMI, ethnicity, marital status, educational level, work status, physical work load, leisure time activity, smoking habits, past medical history including previous surgery for disc herniation and duration of back pain)**	Reported at baseline
– **Blinding questionnaire**	Reported at 3 and 12 months

– **Number of days the last week the patients took the study medication (0-7)**	Reported every week during treatment period by the patient
– **Counts of returned capsules at end of treatment period**	Reported at 3 months by care providers
– **Haematological parameters (White cell counts, thrombocytes, haemoglobin (Hb) and hematocrit (Ht)), measures of kidney (creatinine), liver function (AST / ALT), CRP and Glucose**	assessed and registered monthly during the intervention period and at one year follow up
– **Clinical evaluation (blood pressure, pulse, auscultation of hearth and lunges)**	assessed and registered monthly during the intervention period and at one year follow up
– **Expectations of treatment effect** – **Q: "What is your personal expectation for antibiotic treatment improving your back pain?" Alternatives: "My back pain will be cured" "It results in a clear improvement" "It results in a small improvement" "It results in no improvement" "Don't know"**	Baseline

Data completeness were encouraged by SMS and assessed weekly for all participants. In case of missing data, clinicians and/or patients were contacted to detect the cause.

Results

Figure S2 - Appendix to flowchart

174 patients were listed in the "Other inclusion/exclusion criteria not fulfilled" in Figure 1. The reasons are as follows:

Do not wish to participate in study	53
Modic Changes not present	18
Allergy to penicillin	10
Elevated creatinine or AST/ALT	3
Former lumbar disc surgery, but <12 months since operation	1
Former lumbar disc surgery, but Modic Change at different level	1
Reservations to eat product from swine (gelatin)	2
Regular use of glucocorticoids	1
Regular use of opioids (except codeine and tramadol)	5
Poor language skills (Norwegian)	5
Low compliance	15
Antibiotic treatment within 1 month of study participation	7
Other causes not listed as specific exclusion criteria (such as too old MRI scans, spontaneous improvement in condition, massive comorbidity, seeking commercial treatment in Denmark, low back pain not present, pregnancy in planning, planned elective surgery)	58

The sum of these patients equals 179 (not 174). This is probably due to five double registrations and/or multiple inclusion/exclusion criteria not fulfilled in one or more of the study centers.

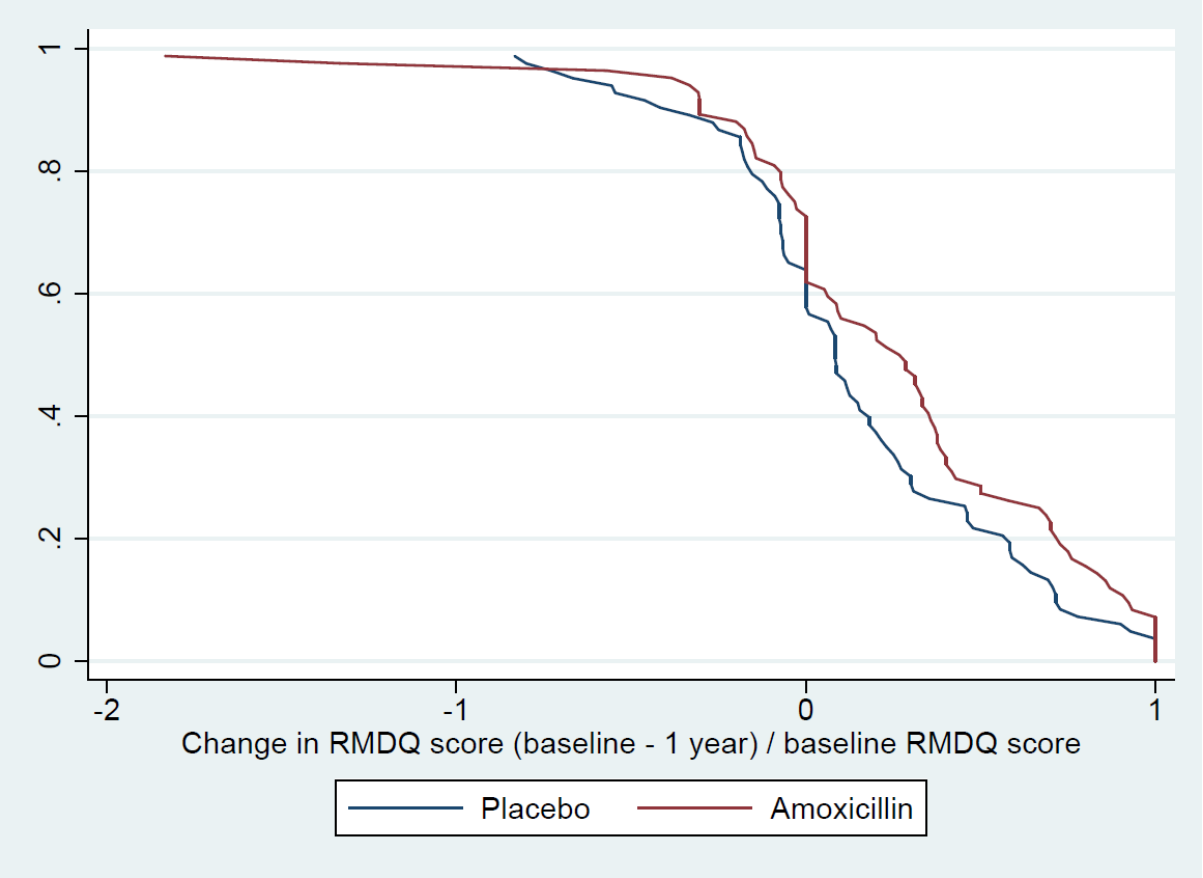
Table S2 - Responder analyses

Improvement in RMDQ score from baseline to 1 year	Treatment groups		Treatment comparison		
	Amoxicillin group (n= 89)	Placebo group (n= 91)	N*	Numbers needed to treat (NNT)	P-value
Improved >30% –no./total no. (%)	40/84 (48%)	24/83 (29%)	167	5.3 (3.0 to 24)	0.01
Improved >50% –no./total no. (%)	23/84 (27%)	18/83 (22%)	167	18 (5.3 to −14)	0.39
Improved >75% –no./total no. (%)	15/84 (19%)	7/83 (8%)	167	10.6 (5.1 to −140)	0.07

Table shows number of patients who improved (>30%, >50% or >75%) from baseline to one year, out of total number who answered RMDQ at baseline and one year. No imputations were performed in Responder analyses

* Number of cases included in the analyses

Figure S3 - Cumulative distribution function of responders

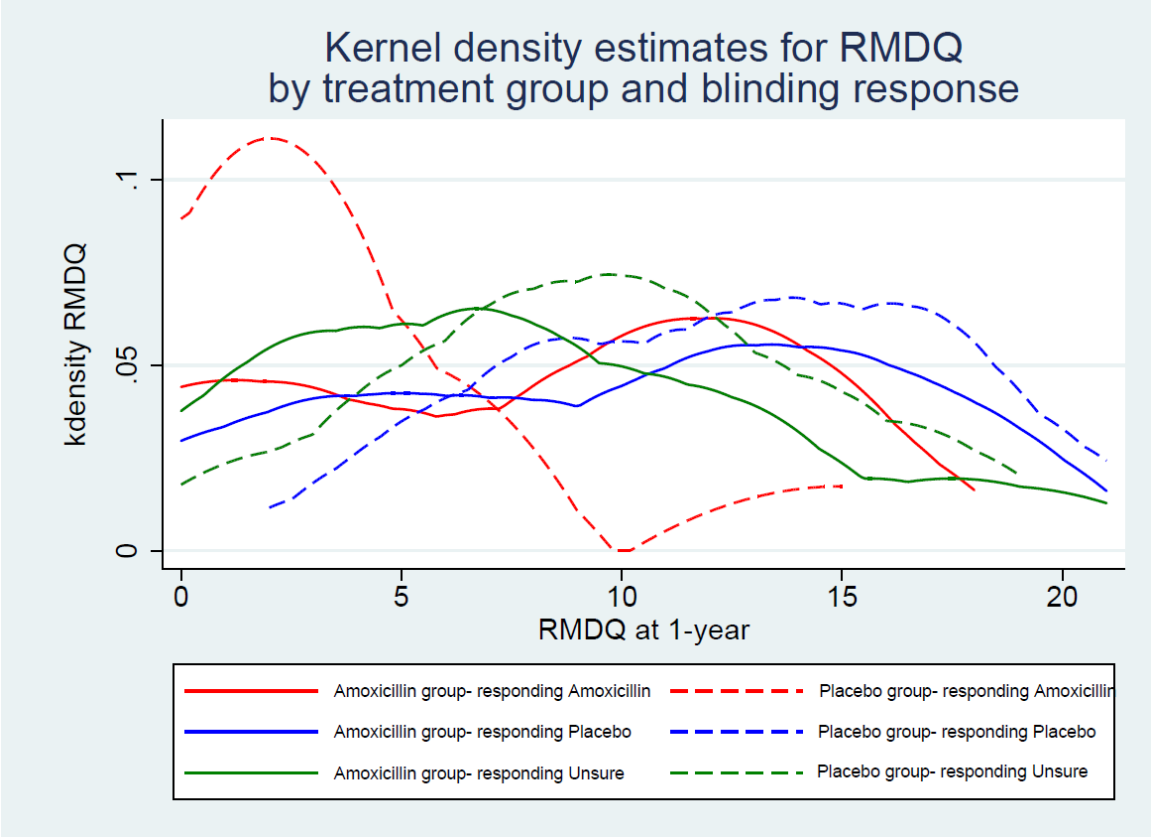


Cumulative proportion of the patients depending on change in the primary outcome (RMDQ) from baseline to one year.

Table S3 - Response to blinding question at one year

	Amoxicillin group (n= 89)	Placebo group (n= 91)
Antibiotics	23	9
Placebo	36	50
Unsure - antibiotic	11	7
Unsure - placebo	12	15
Unsure – missing subgroup	2	2
Missing	5	8

Figure S4 - Kernel density estimates



Smoothed curves of the probability distribution for the number of patients of the RMDQ score at one year (kernel estimator with bandwidth *2.0–2.8*).

The curves allow comparison of the estimated probability density functions of the RMDQ score at 1 year between the two treatment groups (amoxicillin versus placebo) within each of three subgroups. Results were inconsistent. In the subgroup reporting they were unsure what treatment they received, the green curves suggest a slightly higher probability of low RMDQ scores at 1 year in the amoxicillin treatment group (n=25, bold green curve) versus the placebo treatment group (n=24, dotted green curve). Similar minor differences were found in the subgroup reporting they thought they received placebo (amoxicillin treatment group n=36, bold blue curve; placebo treatment group n=50, dottet blue curve). Among those reporting they thought they received amoxicillin, numbers were smaller and results were opposite: the red curves suggest a lower (not higher) probability of low RMDQ score at 1 year in the

amoxicillin treatment group (n=23, bold red line) versus placebo treatment group (n=9, dotted red line). Patients perceiving the received placebo (blue) were generally more likely than the others (green and red) to report higher (worse) RMDQ score at 1 year.

Table S4 – Response to blinding question at one year follow-up.

	Odds Ratio	95% CI	P-value
Amoxicillin group	2.3	0.9 to 5.5	0.07
Responder at 3 months (>30% improvement from baseline)	3.1	1.3 to 7.3	0.01
Drug-related adverse events (any grade)	2.4	1.02 to 5.6	0.045
Increment (placebo group, non-responder and no adverse reaction)	0.029	0.006 to 0.136	<0.001

Odds ratios (CI, p-value) for treatment vs. (placebo or unsure) as response to the blinding question at one year follow-up in the multivariate logistic model.

Table S5 - Sensitivity analyses

Analysis	Treatment groups				Treatment comparison (adjusted mean difference)		
	Amoxicillin		Placebo		ANCOVA		
	N	Mean±SD	N	Mean±SD	N*	Mean (95%CI)	P-Value
RMDQ 1-year							
Per-protocol population	77	9.2 ±6.3	78	10.5 ±5.5	155	−1.4 (-3.0 to 0.2)	0.09
ITT population adjusted for study centre	85	9.0±6.2	84	10.7±5.6	180	−1.7 (-3.2 to −0.1)	0.03
ITT population adjusted for baseline differences	85	9.0±6.2	84	10.7±5.6	180	−1.5 (-3.0 to 0.05)	0.06
Expectation of treatment effect†	62	9.4±6.2	62	11.2±5.6	132	−2.0 (-3.8 to −0.1)	0.04
No expectation of treatment effect†	23	8.0±6.2	22	9.2±5.3	48	0.5 (-2.5 to 3.6)	0.73

* Number of cases included in the analyses after multiple imputations

† Post-hoc analyses, not described in the registry or the Statistical analysis plan. Expectation of treatment effect was defined as “My back pain will be cured”, or “It results in a clear improvement”, no expectation of treatment effect was defined as “It results in a small improvement”, “It results in no improvement”, or “Don’t know” in the response to the question on expectation of treatment effect (table S3)

Table S6 - Primary and secondary outcomes for the separate MC types

	Treatment groups				Treatment comparison (adjusted mean difference)		
	Amoxicillin		Placebo		ANCOVA		
Outcome	N*	Mean±SD	N*	Mean±SD	N¶	Mean (95%CI)	p-value
Type I Modic changes							
RMDQ (0–24)							
Baseline	58	12.9±4.3	60	12.3±3.7	-	-	-
12 months	55	8.2±6.0	56	10.3±5.4	118	−2.3 (-4.2 to −0.4)	0.02
ODI (0–100)†							
Baseline	58	31.3±11.5	58	30.4±9.9	-	-	-
12 months	55	23.4±15.2	56	27.7±13.8	117‡	−5.1 (−9.3 to −0.8)§	0.02
Back pain intensity (0–10) †							
Baseline	58	6.5±1.1	59	6.3±1.3	-	-	-
12 months	55	5.2±2.2	55	4.5±2.5	117‡	−0.8 (−1.6 to 0.0)§	0.06
EQ-5D (−0.59 – 1)†							
Baseline	58	0.55±0.18	60	0.56±0.16	-	-	-
12 months	55	0.66±0.21	55	0.60±0.21	118	0.07 (0.01 to 0.14)§	0.03
Type II Modic changes							
RMDQ (0–24)							
Baseline	30	12.3±5.5	30	13.7±3.5	-	-	-
12 months	30	10.5±6.5	28	11.4±5.9	62	−0.1 (−2.7 to 2.6)	0.95
ODI (0–100)†							
Baseline	30	33.0±11.4	31	34.4±10.6	-	-	-
12 months	30	26.1±14.7	28	31.3±14.3	62	−4.5 (−10.6 to 1.6)§	0.14
Back pain intensity (0–10) †							
Baseline	30	6.3±1.3	31	6.2±1.9	-	-	-
12 months	30	5.0±1.9	28	5.2±2.5	62	−0.3 (−1.3 to 0.7)§	0.52
EQ-5D (−0.59 – 1)†							
Baseline	31	0.53±0.21	31	0.51±0.20	-	-	-
12 months	29	0.61±0.24	28	0.55±0.25	62	0.06 (−0.04 to 0.16)§	0.22

RMDQ Roland-Morris Disability Questionnaire. Score from 0 to 24. Higher scores indicate more severe pain and disability.
ODI Oswestry Disability Index. Score from 0 to 100. Higher scores indicate more severe pain and disability.
EQ5D EuroQol's health-related quality of life. Score from −0.59 to 1. Higher scores indicating better quality of life.
* Number of answered questionnaires of outcome
† Post-hoc analyses, not described in the registry or the Statistical analysis plan
‡ One patient excluded from analysis as the imputation model did not manage to impute all missing variables
§ Estimates smaller than the recommended thresholds for clinical important change within groups (ODI 13 to 20; LBP intensity NRS 2 to 3; Leg pain intensity NRS 2 to 3.5; EQ5D 0.11 to 0.30)
¶ Number of cases included in the analyses after multiple imputations

Figure S5 - Intensity grade 1 adverse events, frequency distribution

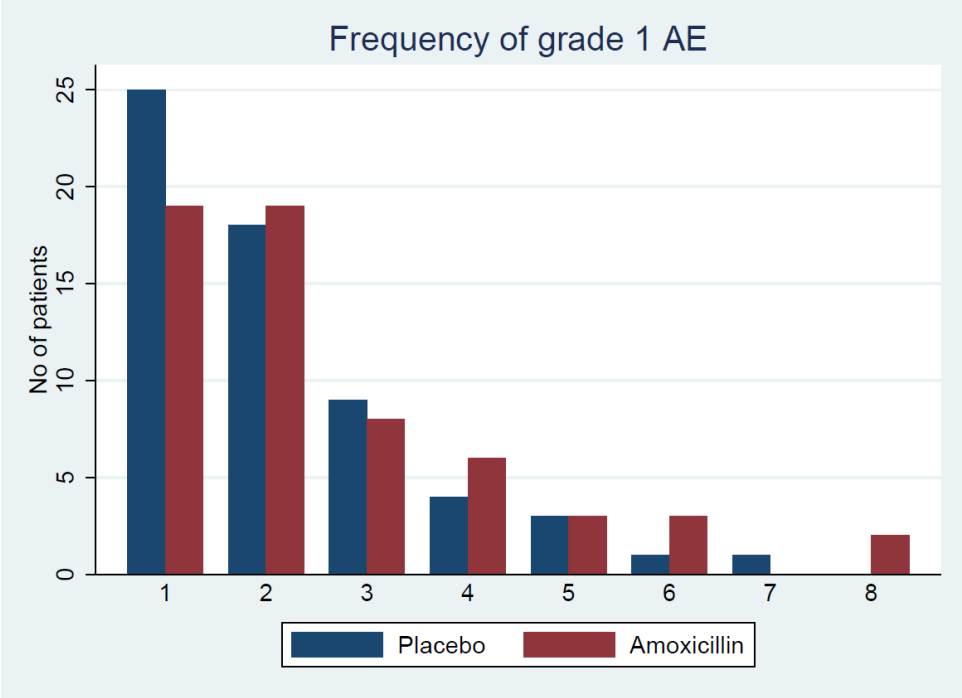


Figure S6 - Intensity grade 2 adverse events, frequency distribution

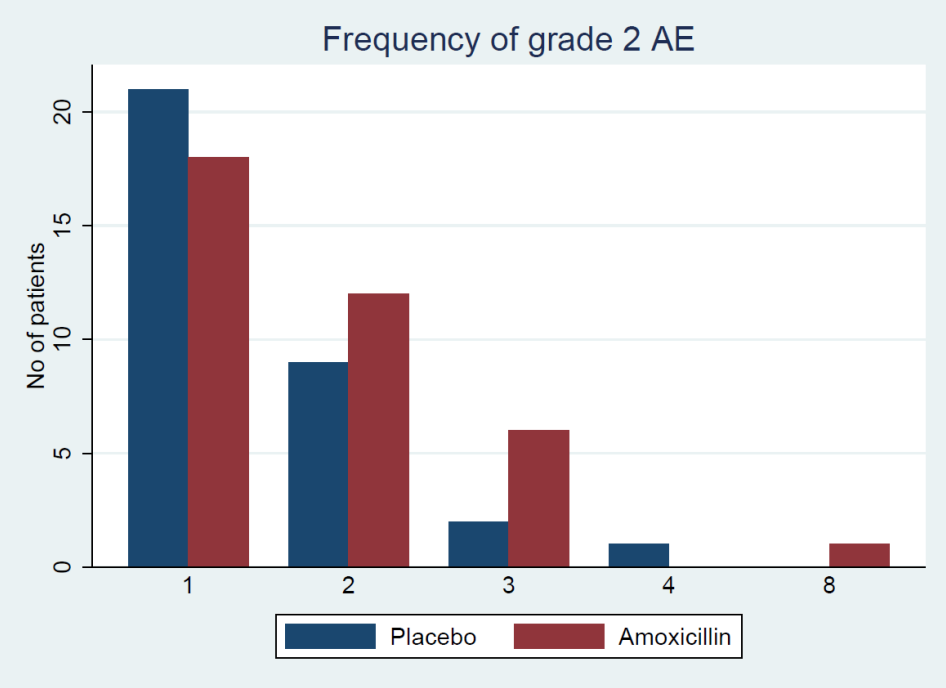


Figure S7 - Serious adverse events or intensity grade 3-4 adverse events, frequency distribution

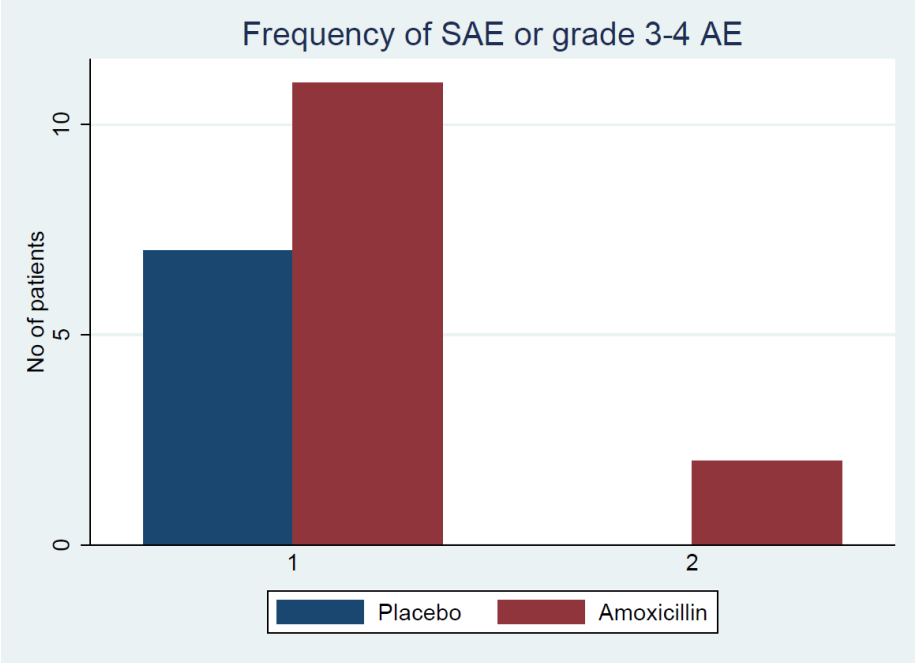


Table S7 - Most frequent adverse event types by attribution group and treatment group

Adverse event type	Placebo group		Amoxicillin group		Total
	Unrelated/ Unlikely/NA	Drug related	Unrelated/ Unlikely/NA	Drug related	
Diarrhoea	6	14	1	23	44
Back pain	20	1	13	0	34
Influenza like illness	6	0	9	1	16
Headache	8	1	4	1	14
Nausea	0	3	0	9	12
Nasopharyngitis	6	0	4	0	10
Hepatic enzyme increased	2	1	4	3	10
Abdominal pain upper	0	2	1	6	9
Flatulence	1	3	0	4	8
Fatigue	0	2	2	3	7
Influenza	3	0	4	0	7
Cardiac murmur	4	0	2	0	6
Gastroenteritis	4	0	2	0	6
Rash maculo-papular	2	1	1	2	6
Arthralgia	3	0	2	0	5
Dry mouth	0	2	2	1	5
General physical health deterioration	3	0	2	0	5
Pain in extremity	3	0	2	0	5
Rash	0	0	1	4	5
Tooth discoloration	0	0	3	2	5
Urinary tract infection	1	0	3	1	5
Abdominal discomfort	0	2	0	2	4
Cough	2	0	2	0	4
Dental plaque	1	0	1	2	4
Dizziness	0	2	2	0	4
Dyspnoea	3	0	0	1	4
Hypertension	2	0	2	0	4
Vulvovaginal candidiasis	0	0	0	4	4
Vulvovaginitis	0	0	0	4	4
Chest injury	2	0	1	0	3
Conjunctivitis	3	0	0	0	3
Constipation	1	2	0	0	3
Dry skin	0	0	2	1	3
Fall	1	0	2	0	3
Fungal infection	0	0	0	3	3
Gastrooesophageal reflux disease	0	0	0	3	3
Leukocytosis	3	0	0	0	3
Musculoskeletal pain	0	0	3	0	3
Oesophagitis	0	0	2	1	3
Sinusitis	1	0	2	0	3
Urticaria	0	0	0	3	3

Terminology compliant with CTCAE

References:

1. Hayden JA, Chou R, Hogg-Johnson S, Bombardier C. Systematic reviews of low back pain prognosis had variable methods and results—guidance for future prognosis reviews. Journal of Clinical Epidemiology 2009;62:781-96.e1.
2. Bråten LC RM, Espeland A, Storheim K, Zwart JA, Hellum C, Aßmus J. 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.gov2018.
3. Hernán MA, Robins JM. Per-Protocol Analyses of Pragmatic Trials. [S.l.]2017:1391-8.
4. Grotle M, Vøllestad M, Brox M. Cross-cultural adaptation of the Norwegian versions of the Roland-Morris Disability Questionnaire and the Oswestry Disability Index. Journal of Rehabilitation Medicine 2003;35:241-7.
5. Chiarotto A, Boers M, Deyo RA, et al. A CORE OUTCOME MEASUREMENT SET FOR LOW BACK PAIN CLINICAL TRIALS. Orthopaedic Proceedings 2018;100-B:29-.
6. Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000;25:2940.
7. Dworkin HR, Turk CD, Farrar TJ, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9-19.
8. EuroQol - a new facility for the measurement of health-related quality of life. Health policy 1990;16:199-208.
9. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. Behav Sci 1974;19:1-15.
10. Waddell G, Newton M, Henderson I, Somerville D, Main CJ. A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. Pain 1993;52:157-68.
11. Eriksen HR, Ihlebaek C, Ursin H. A scoring system for subjective health complaints (SHC). Scandinavian journal of public health 1999;27:63.