Performance of a Deep Learning Algorithm to Detect Positive SIJ MRI
According to the ASAS Definition in axSpA Patients

Objective
To assess the ability of a previously trained deep learning algorithm to identify the presence of sacroiliac joint inflammation in MRI scans in a study cohort of patients with axial spondyloarthritis.

Background
• MRI of the sacroiliac joints (SIJ) is an essential tool in the clinical diagnosis of patients with axial spondyloarthritis (axSpA), but it involves high knowledge of characteristic MR features, their definitions and reliability of identification among general radiologists and rheumatologists.1

• A trained deep learning algorithm to detect the presence of inflammation in SIJ MRI scans has previously been developed with promising results in a small patient cohort.2

• Further evaluation of the deep learning algorithm in larger external validation cohorts, specifically in non-radiographic and non-spondyloarthritis axial inflammatory populations, is required to assess its potential for peri-clinical use.

Methods
MRA Scans
• Baseline SIJ MRI scans were collected from patients with nr-axSpA or nr-axSpA in two prospective randomized controlled trials cohorts, RAPID-axSpA (NCT01087762) and C-OPTIMISE (NCT02505542).2,3

• The MRI scans were centrally evaluated by two human experts, and an adjudicator in case of disagreement, for the presence of SIJ inflammation as defined by the 2009 Assessment in Ankylosing Spondylitis International Society (ASAS) definition of MRI positivity (MRI+).1

• The scans were then processed by the previously trained deep learning algorithm,1 blinded to clinical information and central expert readings.

Model Performance Evaluation
• The agreement between the deep learning algorithm and expert reader for the binary classification of MRI+ SIJ scans (MRI+ vs MRI−) was assessed using sensitivity, specificity, positive prediction value (PPV), negative prediction value (NPV), absolute agreement and Cohen’s kappa.

• Bootstrapping was used to construct 95% confidence intervals (CIs).

Results
Baseline MRA Scans and Patient Characteristics
• In total, 731 MRI SIJ scans were collected from pooled patients in RAPID-axSpA (n=729) and C-OPTIMISE (n=7), comparing the validation set (Figure 2A).

• In the pooled study population, 44% (n=326) were patients with nr-axSpA and 59.6% (n=436) were MRI+ as determined by expert readings (Table 1).

Model Validation
• Comparing the trained algorithm with the central expert readings for the classification of MRI+ vs MRI− scans on the pooled validation set yielded a sensitivity of 0.79 (95% CI: 0.70–0.85), specificity of 0.76 (95% CI: 0.70–0.83), PPV of 0.54 (95% CI: 0.46–0.60) and absolute agreement of 0.74 (95% CI: 0.62–0.80) (Table 1).

Conclusions
This study is the first evaluation of a deep learning algorithm to detect SIJ MRI positivity.2-6

This study also provides a rigorous evaluation of the algorithm’s performance in a larger external validation set of patients with axSpA from two clinical trials.

This suggests that a detection algorithm for SIJ MRI has the potential to support clinicians in the diagnosis of patients with axSpA.

Table 1

<table>
<thead>
<tr>
<th>Metric</th>
<th>Algorithm MRI+</th>
<th>Expert MRI+</th>
<th>Algorithm MRI–</th>
<th>Expert MRI–</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.79 (0.70–0.85)</td>
<td>0.79 (0.70–0.85)</td>
<td>0.76 (0.70–0.83)</td>
<td>0.76 (0.70–0.83)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.76 (0.70–0.83)</td>
<td>0.76 (0.70–0.83)</td>
<td>0.77 (0.70–0.85)</td>
<td>0.77 (0.70–0.85)</td>
</tr>
<tr>
<td>PPV</td>
<td>0.54 (0.46–0.60)</td>
<td>0.54 (0.46–0.60)</td>
<td>0.50 (0.41–0.58)</td>
<td>0.50 (0.41–0.58)</td>
</tr>
<tr>
<td>NPV</td>
<td>0.92 (0.87–0.95)</td>
<td>0.92 (0.87–0.95)</td>
<td>0.84 (0.78–0.88)</td>
<td>0.84 (0.78–0.88)</td>
</tr>
</tbody>
</table>

a Data for the pooled validation set (N=731).

Figure 1
Study flow diagram

Figure 2
Performance results comparing the deep learning algorithm and human experts for classification of SIJ MRI scans