Recruitment procedure to maximise inclusion of progressors in OA clinical studies using subjects from existing cohorts

**APPRAOCH cohort**
- new longitudinal cohort focused on fast knee OA progressors
- recruited from 5 existing European OA cohorts using machine learning models trained to estimate progression probability

<table>
<thead>
<tr>
<th>KL grade</th>
<th>JSN grade</th>
<th>JSW</th>
<th>follow-up</th>
<th>active years</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHECK</td>
<td>✓</td>
<td>✓</td>
<td>1–10 years</td>
<td>2006–2016</td>
</tr>
<tr>
<td>MUST</td>
<td>✓</td>
<td>✓</td>
<td>2010–2015</td>
<td></td>
</tr>
<tr>
<td>HOSTAS</td>
<td>✓</td>
<td>✓</td>
<td>2009–2007</td>
<td></td>
</tr>
<tr>
<td>DIGICOD</td>
<td>1 year</td>
<td>✓</td>
<td>2013–2017</td>
<td></td>
</tr>
<tr>
<td>PROCOC</td>
<td>✓</td>
<td>✓</td>
<td>1991–2016</td>
<td></td>
</tr>
</tbody>
</table>

**Machine learning from historical data**
- builds computational models from sample inputs, and uses them to predict outcomes

**Input data**
- includes patients data
- categorises by pain, no progression, structure

**Output data**
- learning categories
- predicted categories
- pain structure
- no progression

**Machine learning algorithm**
- input data to predict outcomes

**Two-stage recruitment process**
- new patients to screening
- selection criteria
- screening ML model
- approach cohort

**INPUT LEARNING OUTPUT**

**Input categories**
- structure periods
  - minimum total JSW, must decrease by at least 0.3mm per year

- pain periods
  - must experience progressive or intense sustained pain
  - pain increase must be at least 5 WOMAC points per year
  - pain at the follow-up must be significant (≥ 40 WOMAC points)
  - special exception: rapid pain progression (≥ 10 points per year)

**Data harmonisation**
- transformation of CHECK data (common subset of attributes)
- attributes mapping to CHECK

**Measurables**
- basic: age, sex, BMI
- pain intensity: KOOS, NRS
- KDA: bone density, eminence height, joint space width, varus angle, osteophyte area

**Ranking based on model confidence**
- the ML model is composed of two sub-models, separately predicting pain (P) and structure (S) related progression
- the outcome of the model is per subject probability of becoming an OA progressor
- top probabilities are used to rank the subjects
- top subjects in the ranking are more likely to progress

<table>
<thead>
<tr>
<th>P</th>
<th>S</th>
<th>Score</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3</td>
<td>4.0</td>
<td>0.707</td>
<td>0.805</td>
</tr>
<tr>
<td>3.7</td>
<td>3.7</td>
<td>0.707</td>
<td>0.805</td>
</tr>
<tr>
<td>3.3</td>
<td>3.3</td>
<td>0.707</td>
<td>0.805</td>
</tr>
<tr>
<td>2.1</td>
<td>2.1</td>
<td>0.707</td>
<td>0.805</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>0.707</td>
<td>0.805</td>
</tr>
</tbody>
</table>

**APPRAOCH lifespan**
- last visit
- baseline shift
- progression

**Models training**
- using harmonised CHECK periods
- baseline shift of 0, 2, 3, and 5 years

**APPRAOCH cohort**
- 2–4 years 2009–2017
- ✓ ✓

**APPRAOCH lifespan**
- 2014
- 2017
- 2019

**Model training**
- 3 years shift
- 5 years shift

**Patient data**
- prediction
- ranking

**Multi-model prediction**
- using harmonised CHECK periods
- baseline shift of 0, 2, 3, and 5 years

**Selection criteria**
- new patients
- screening ML model
- approach cohort

**Screening ML model**
- selection criteria
- screening ML model
- approach cohort

**Procedure**
- transformation of CHECK data (common subset of attributes)
- attributes mapping to CHECK

**Structure periods**
- 3 years shift
- 5 years shift

**Subjects data**
- symptoms
- procedures
- variables

**Patients data**
- symptoms
- procedures
- variables

**Two-stage recruitment process**
- new patients
- selection criteria
- screening ML model
- approach cohort

**Input categories**
- structure periods
- pain periods

**Output categories**
- structure periods
- pain periods

**Pain periods**
- 3 years shift
- 5 years shift

**Subjects data**
- symptoms
- procedures
- variables

**Selection criteria**
- new patients
- screening ML model
- approach cohort

**Screening ML model**
- selection criteria
- screening ML model
- approach cohort

**Two-stage recruitment process**
- new patients
- selection criteria
- screening ML model
- approach cohort

**Input data**
- inclusion
- patient
- periods

**Output data**
- prediction
- ranking

**Input categories**
- minimum total JSW, must decrease by at least 0.3mm per year

**Pain periods**
- must experience progressive or intense sustained pain
- pain increase must be at least 5 WOMAC points per year
- pain at the follow-up must be significant (≥ 40 WOMAC points)
- special exception: rapid pain progression (≥ 10 points per year)