* Scraping recommended, not recommended (classify no recommendation as not recommended)
  + 3 categories.: recommended with condition, recommended and not recommended
* Only download all committee papers for now

**2 Options:**

* feature extraction using NLP (good for cv) + log regression (method 1)
* No Feature extraction completely use a deep learning model for decision, with some kind of interpretability (method 2)

Different model have different probabilities

**Next steps:**

* Each member to explore method 1 using a ‘test’ set. Understand where the different features lies, and will there be a way to extract them (using NLP)

### Features for consideration

#### Basic information:

country of HTA

year of application (maybe not necessary)

Type of disease

* Cardiovascular
* Chronic
* Nervous System
* Obstetrics/Gynaecology
* Urinary-Track Diseases
* Cancer
* Rare Disease (Orphan Drugs)
* Ultra-Rare DIsease (Ultra Orphan Drugs)
* Musculoskeletal diseases

#### Advanced information:

innovation

Second submission?

Disease severity

QALY benefits

#### Clinical trials details:

The paediatric population (test population between 0-18)

Clinical uncertainties

Adverse reactions

How many RCTs

Quality of evidence

number of patients

treat unmet need

Observational studies

#### Economic:

cost-effectiveness

ICER (Incremental Cost-Effectiveness Ratio)