Supplementary material 1

Sample Size

Power is based on the primary outcome stipulated by the HTA-diagnostic accuracy for Crohn's disease extent. In this section, MRI will be used as the basis of statistical sample size for MRE.

There are two aspects to correctly assigning disease extent-correctly detecting the presence of disease AND correctly assigning its segmental location. For example, a test which correctly identifies disease in the terminal ileum of the small bowel but misses disease in the proximal bowel (e.g. jejunum) will likely result in an incorrect patient management decision i.e. such a test would be inaccurate for defining the extent of Crohn's disease. Power is thus based on a two facetted compound accuracy measure (disease presence and disease location).

Patients with disease identified by reference test

Primary outcome: Test accurate for disease extent	Correct identification of disease presence	Test accurate for disease extent?
Y	Yes –disease identified	Yes-all segments identified
N	Yes –disease identified	No- one or more segments missed
N	Yes –disease identified	No- incorrect segment(s) identified
N	N- no disease identified	No disease identified

Power calculation

Comparison of MRI to USS accuracy - both against a composite reference standard

Study powered to show difference in sensitivity for disease extent (compound of disease presence and correct disease location). Assume moderate correlation between the tests as both are imaging tests:

Paired study design - all tests on all patients

90% power type II error, type I error 5% ($p<0.05$)

Per patient unit of analysis

Combined patient population (i) new patients diagnosed with Crohn's (ii) patients with suspected relapse. Both groups have approximately 70% prevalence of small bowel disease

Assumptions

Sensitivity for correct disease presence (see assumption 1 below)

MRI 93%

USS 88%

Sensitivity for correct disease location (see assumption 2 below)

MRI 90%

USS 83% (encompassing 30% sensitivity for the 5-10% of patients with proximal small bowel disease)

Compound accuracy measure (disease presence and disease location)

10% difference in sensitivity between tests

MRI **83%** = 93% (disease presence) x 90% (disease location)

USS **73%** = 88% (disease presence) x 83% (disease location)

68% test results are positive with both USS and MRI. Reasonable correlation assumed as both tests are imaging tests. A higher correlation would result in a lower sample size.

Prevalence of small bowel Crohn's disease -**70%** (new diagnosis and relapsing patients). See assumption 3 below

Sample size calculation

Sample size method[1]

Power beta	Alpha	Sens MRI	Sens US	% Patients US+MRI+	% Discrepant cells	Total DP	Prevalence	Total N	Total with 10% LFU*
90%	0.05	0.83	0.73	0.68	0.2	210	0.7	301	334
80%	0.05	0.83	0.73	0.68	0.2	157	0.7	224	249

Total cohort=301 (210 patients with disease)

Allowing 10% loss to follow up, **total cohort=334** (167 new diagnosis patients and 167 relapse patients)

Evidence base for underlying assumptions of test diagnostic accuracy

Assumption 1: Estimates for sensitivity of disease detection with USS and MRI

Summary sensitivity for detection of small bowel disease:

MRI-93%, USS-88%[2]

Assumption 2: Disease location

The trial team have contacted the authors of this systematic review and there is insufficient data to look at the differential sensitivity of imaging tests for proximal and distal small bowel disease.

Diagnostic accuracy for proximal small bowel disease[3]

Disease Prevalence assumptions[4]

Assumption 3: Disease presence

Highest level of evidence is a systematic review[2]

Study power-Secondary outcomes

Disease activity

Methods

Crohn's disease activity will be considered on a per segment (for the terminal ileum) and per patient basis.

Comparison of MRI to USS accuracy - both against a composite reference standard

Study powered to show a difference in sensitivity for activity. Assume moderate correlation between imaging tests.

Paired study design - all tests on all patients

80% power type II error, type I error 5% ($p<0.05$)

Combined Patient population (i) new patients diagnosed with Crohn's (ii) patients with suspected relapse. Both groups have approximately 70% prevalence of small bowel disease

Sample size method

Per segment (terminal ileum)

Segmental assessment of disease activity can only be meaningfully acquired using an endoscopic reference (global markers such as HBI, calprotectin are not segment specific). The terminal ileum is the most robust segment to acquire endoscopic assessment of disease activity given its ease of identification and fundamental importance in the diagnosis and assessment of Crohn's disease. Endoscopic evaluation of the terminal ileum will be available in around 200 patients (all new

diagnosis and one third of relapse). The HTA requirement to study those with a new diagnosis of Crohn's disease means prospective collection of CDEIS will not be possible i.e. endoscopy will in the main be performed before recruitment and CDEIS is not recorded as part of routine clinical practice). Activity in the terminal ileum will thus be assigned by the consensus reference panel based on the endoscopic report, endoscopic images (photographic documentation of the terminal ileal appearances is routine at recruitment sites), and histology of TI biopsies, also routine

Assumptions

Sensitivity for correct presence of active disease (see assumption 4,5 below)

MRI 75%

USS 60%

50% test results are positive with both USS and MRI. Reasonable correlation assumed as both tests are imaging tests. A higher correlation would result in a lower sample size.

Prevalence of small bowel Crohn's disease is **70%** (new diagnosis and relapsing patients). See assumption 3 above.

One segment per patient: terminal ileum

Power	Type I error	Sens MRI	Sens US	% Patients US+MRI+	% Discrepant cells	total DP	prevalence	Total N	Total with 10% LFU
80%	0.05	0.75	0.60	0.50	0.35	122	0.7	175	195

Sample size calculation

Sample size method[1]

Total N=122 disease positive segments at one per patient. This corresponds to 175 patients at 70% per patient prevalence and 80% power. **195 patients** will be required allowing 10% loss to follow up.

Endoscopic evaluation of the terminal ileum will be available in around 200 patients (all new diagnosis (n=167) and one third of relapse (n=55; 0.33x167).

Per patient

Sensitivity MRI 88% (see assumption 6)

with cohort powered for primary outcome, we have 80% power to detect a 10% change in activity per patient.

Power	Alpha	Sens MRI	Sens US	% Patients US+MRI+	% Discrepant cells	total DP	prevalence	Total N	Total with 10% LFU
80%	0.05	0.88	0.78	0.70	0.26	204	0.7	292	324

Evidence base for underlying assumptions of test classification of disease activity:

Assumption 4: Classification of activity per segment

Two systematic reviews include meta-analyses of MRI in the classification of Crohn's disease activity[5, 6]

The largest study directly comparing USS with MRI in the same patients include 30 patients, 23 with disease[7]

Assumption 5: Per segment sensitivity for correct disease activity classification (encompassing prevalence estimates)[5].

MRI-78%

USS-60%

I.e. an **18%** difference between tests

Assumption 6: Per patient sensitivity for disease activity

An assumption of 88% sensitivity for MRI is based on 6 studies with a total of 118 DP patients (range 7 to 28 per study)[5]. Although the Panes SR[5] identifies a range of sensitivity for USS of 77-100% with a summary of 85% sensitivity this is based on 5 studies with between 23 and 47 patients with active disease per study. However these results are likely to be over optimistic due to several sources of bias

(1) threshold effects: sensitivity is quoted for two threshold values for bowel wall thickness, >2.5mm for ileal segments >3.0mm for all segments, with sensitivity of **75% and 48%** respectively. If thresholds are chosen to optimise diagnostic performance within a study, sensitivity values are over-estimated.

(2) disease spectrum bias: sensitivity varies from **33% to 67%** depending on the segment with active disease and a threshold of >3.0mm. In addition the sensitivity varies with disease severity (mild, moderate, severe).

Studies with very small numbers of patients will have high potential for disease spectrum bias

References

1. Taylor S, Mallett S, Bhatnagar G, Bloom S, Gupta A, Halligan S, *et al.* METRIC (MREnterography or ulTRasound in Crohn's disease): a study protocol for a multicentre, non-randomised, single-arm, prospective comparison study of magnetic resonance enterography and small bowel ultrasound compared to a reference standard in those aged 16 and over. *BMC Gastroenterol* 2014;**14**:142. http://dx.doi.org/10.1186/1471-230x-14-142

2. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med* 2002;**21**:835-52.

3. Horsthuis K, Bipat S, Bennink RJ, Stoker J. Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: meta-analysis of prospective studies. *Radiology* 2008;**247**:64-79. http://dx.doi.org/10.1148/radiol.2471070611

4. Fraquelli M, Colli A, Casazza G, Paggi S, Colucci A, Massironi S, *et al.* Role of US in detection of Crohn disease: meta-analysis. *Radiology* 2005;**236**:95-101. http://dx.doi.org/10.1148/radiol.2361040799

5. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007;**369**:1641-57. http://dx.doi.org/10.1016/s0140-6736(07)60751-x

6. Panes J, Bouzas R, Chaparro M, Garcia-Sanchez V, Gisbert JP, Martinez de Guerenu B, *et al.* Systematic review: the use of ultrasonography, computed tomography and magnetic resonance imaging for the diagnosis, assessment of activity and abdominal complications of Crohn's disease. *Aliment Pharmacol Ther* 2011;**34**:125-45. http://dx.doi.org/10.1111/j.1365-2036.2011.04710.x

7. Horsthuis K, Bipat S, Stokkers PC, Stoker J. Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. *Eur Radiol* 2009;**19**:1450-60. http://dx.doi.org/10.1007/s00330-008-1287-0

8. Miao YM, Koh DM, Amin Z, Healy JC, Chinn RJ, Zeegen R, *et al.* Ultrasound and magnetic resonance imaging assessmentof active bowel segments in Crohn's disease. *Clin Radiol* 2002;**57**:913-8.