## Supplementary information 4. Quality assessment

### Prognostic accuracy and clinical impact

#### IBDX

Table 1. Harrell 2010

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Harrell 2010 | | |
| Study details (journal, year, volume, page range) | *Gastroenterology*, 2010, 138, (5), S529 | | |
| Type of report (full paper/only abstract/conference abstract) | Conference abstract | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes. | Unclear |
| Description of the source population or population of interest | No, insufficient detail in abstract. |
| Description of the baseline study sample | No, insufficient detail in abstract. |
| Adequate description of the sampling frame and recruitment | No, insufficient detail in abstract. |
| Adequate description of the period and place of recruitment | No, insufficient detail in abstract. |
| Adequate description of inclusion and exclusion criteria | No, insufficient detail in abstract. |
| Study attrition | Adequate response rate for study participants | Yes. | Unclear |
| Description of attempts to collect information on people who dropped out | Unclear. |
| Reasons for loss to follow-up are provided | Unclear. |
| Adequate description of those lost to follow-up | Unclear. |
| There are no important differences between people who completed the study and those who did not | Unclear. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes. | Unclear |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes. |
| Continuous variables are reported or appropriate cut points are used | Unclear. |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Unclear. |
| Adequate proportion of the study sample has complete data for the prognostic factor | Unclear. |
| Appropriate methods of imputation are used for missing prognostic factor data | Unclear. |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes. | Unclear |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Unclear. |
| The method and setting of outcome measurement are the same for all those in the study | Unclear. |
| Study confounding | Most important confounders are measured | Confounders are not discussed in the abstract. Assessment of efficacy of tool is based on event rate of complication of disease (intestinal fistula and/or stricture) and need for surgery. | Unclear |
| Clear definitions of the important confounders measured are provided |
| Measurement of all important confounders is adequately valid and reliable |
| The method and setting of confounding measurement are the same for all those in the study |
| Appropriate methods are used if imputation is used for missing confounder data |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) |
| Important potential confounders are accounted for in the analysis |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | No | Unclear |
| Strategy for model building is appropriate and is based on a conceptual framework or model | Unclear |
| The selected statistical model is adequate for the design of the study | Unclear |
| There is no selective reporting of results | Unclear |
| Abbreviation: IBD, inflammatory bowel disease. | | | |

Table 2. Paul 2015

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Paul 2015 | | |
| Study details (journal, year, volume, page range) | *J Crohns Colitis*, 2015, 9, (6), 445–451 | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes  Study enrolled those with ulcerative colitis and those with Crohn’s disease. Those enrolled had a diagnosis of disease for more than one year. | Low |
| Description of the source population or population of interest | Yes |
| Description of the baseline study sample | Yes |
| Adequate description of the sampling frame and recruitment | Yes |
| Adequate description of the period and place of recruitment | Yes |
| Adequate description of inclusion and exclusion criteria | Yes |
| Study attrition | Adequate response rate for study participants | Yes  Relevant samples collected from all those enrolled. | Low |
| Description of attempts to collect information on people who dropped out | Not applicable. |
| Reasons for loss to follow-up are provided | Not applicable. |
| Adequate description of those lost to follow-up | Not applicable. |
| There are no important differences between people who completed the study and those who did not | Not applicable. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes  Changes in serum levels of individual antibodies. | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes  Authors followed the manufacturer’s instructions to generate unit of measurement. |
| Continuous variables are reported or appropriate cut points are used | Yes. |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes. |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes. |
| Appropriate methods of imputation are used for missing prognostic factor data | Not applicable. |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes for the outcome of interest to the systematic review reported here.  The authors of the review appreciate that it will be difficult to determine the true clinical impact of the tool. | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Analyses of clinical data and serological assessments were carried out in a masked manner without knowledge of patient’s diagnosis and medical history. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | Confounders are not discussed in the full publication | Unclear |
| Clear definitions of the important confounders measured are provided |
| Measurement of all important confounders is adequately valid and reliable |
| The method and setting of confounding measurement are the same for all those in the study |
| Appropriate methods are used if imputation is used for missing confounder data |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) |
| Important potential confounders are accounted for in the analysis |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | No model built. |
| The selected statistical model is adequate for the design of the study | Yes |
| There is no selective reporting of results | Yes |
| Abbreviation: IBD, inflammatory bowel disease. | | | |

Table 3. Rieder 2010b

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Rieder 2010b | | |
| Study details (journal, year, volume, page range) | *Inflamm Bowel Dis*, 2010, 16, 263–274.  Related paper: Rieder 2011a *PLoS ONE*, 2011, 6, (5), e18172 (presents data on subgroup of people reported in Rieder 2010b as a longitudinal analysis).  Related paper: Rieder 2010d *Gastroenterology*, 2010, 138, (5), S522 (conference abstract for Rieder 2011a). | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper. | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes  All people analysed have a diagnosis of CD. However, there is a mixed population in terms of those with a new diagnosis and with an established diagnosis, as well as presence of complicated disease at baseline versus no complications. Data are not reported separately for the various subgroups. | Moderate |
| Description of the source population or population of interest | Yes |
| Description of the baseline study sample | Yes |
| Adequate description of the sampling frame and recruitment | Yes |
| Adequate description of the period and place of recruitment | Yes |
| Adequate description of inclusion and exclusion criteria | Yes |
| Study attrition | Adequate response rate for study participants | Yes  Samples were available for all those enrolled with CD. | Low |
| Description of attempts to collect information on people who dropped out | Not applicable. |
| Reasons for loss to follow-up are provided | Not applicable. |
| Adequate description of those lost to follow-up | Not applicable. |
| There are no important differences between people who completed the study and those who did not | Not applicable. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes |
| Continuous variables are reported or appropriate cut points are used | Yes |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes |
| Appropriate methods of imputation are used for missing prognostic factor data | Not applicable |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Analyses based on stored blood samples. The authors sent the samples to Glycominds for analysis, which was carried out in a masked manner. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | Important baseline characteristics are adjusted for in statistical analyses. Factors adjusted for were age, gender, BMI, disease activity and duration, age at diagnosis, and disease location as potential confounders. The authors comment that it is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery. | Moderate |
| Clear definitions of the important confounders measured are provided | Yes |
| Measurement of all important confounders is adequately valid and reliable | Yes |
| The method and setting of confounding measurement are the same for all those in the study | Yes |
| Appropriate methods are used if imputation is used for missing confounder data | Not applicable |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) | No |
| Important potential confounders are accounted for in the analysis | Yes – see comment above |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | No model built. |
| The selected statistical model is adequate for the design of the study | Yes |
| There is no selective reporting of results | Yes |
| Abbreviation: IBD, inflammatory bowel disease. | | | |

Table 4. Rieder 2010c

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Rieder 2010c | | |
| Study details (journal, year, volume, page range) | *Inflamm Bowel Dis*, 2010, 16, (8), 1367–1375. | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper. | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes  All people analysed have a diagnosis of CD. However, there is a mixed population in terms of those with no complications of disease at baseline, and no prior surgery versus those with complications and/or prior surgery. People had to have at least 3 years of follow-up to be eligible. Subgroup data are reported for subgroups of potential interest. | Low |
| Description of the source population or population of interest | Yes |
| Description of the baseline study sample | Yes |
| Adequate description of the sampling frame and recruitment | Yes |
| Adequate description of the period and place of recruitment | Yes |
| Adequate description of inclusion and exclusion criteria | Yes |
| Study attrition | Adequate response rate for study participants | Yes  Samples were available for all those enrolled with CD. | Low |
| Description of attempts to collect information on people who dropped out | Not applicable. |
| Reasons for loss to follow-up are provided | Not applicable. |
| Adequate description of those lost to follow-up | Not applicable. |
| There are no important differences between people who completed the study and those who did not | Not applicable. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes |
| Continuous variables are reported or appropriate cut points are used | Yes |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes |
| Appropriate methods of imputation are used for missing prognostic factor data | Not applicable |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Analyses based on stored blood samples. The authors sent the samples to Glycominds for analysis, which was carried out in a masked manner. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | The authors comment that it is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery.  The authors proposed age, gender, BMI, disease activity and duration, age at diagnosis, and disease location as potential confounders. | Moderate |
| Clear definitions of the important confounders measured are provided | Yes |
| Measurement of all important confounders is adequately valid and reliable | Yes |
| The method and setting of confounding measurement are the same for all those in the study | Yes |
| Appropriate methods are used if imputation is used for missing confounder data | Not applicable |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) | No |
| Important potential confounders are accounted for in the analysis | Yes  Authors carried out a regression analysis to account for the potential confounders. |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | No model built. |
| The selected statistical model is adequate for the design of the study | Yes |
| There is no selective reporting of results | Yes |
| Abbreviation: IBD, inflammatory bowel disease. | | | |

Table 5. Rieder 2012

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Rieder 2012 | | |
| Study details (journal, year, volume, page range) | *Inflamm Bowel Dis*, 2012, 18, (7), 1221–1231.  Related paper: Rieder 2011b J Crohns Colitis, 2011, 5, (1), S48 (conference abstract). | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper. | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes  All children analysed have a diagnosis of CD. However, there is a mixed population in terms of those with a new diagnosis and with an established diagnosis, as well as presence of complicated disease at baseline versus no complications. Data are not reported separately for the various subgroups. | Moderate |
| Description of the source population or population of interest | Yes |
| Description of the baseline study sample | Yes |
| Adequate description of the sampling frame and recruitment | Yes |
| Adequate description of the period and place of recruitment | Yes |
| Adequate description of inclusion and exclusion criteria | Yes |
| Study attrition | Adequate response rate for study participants | Yes  All people were eligible for analysis and samples were available for all children. | Low |
| Description of attempts to collect information on people who dropped out | Not applicable |
| Reasons for loss to follow-up are provided | Not applicable |
| Adequate description of those lost to follow-up | Not applicable |
| There are no important differences between people who completed the study and those who did not | Not applicable |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes |
| Continuous variables are reported or appropriate cut points are used | Yes |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes |
| Appropriate methods of imputation are used for missing prognostic factor data | Not applicable |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Analyses based on stored blood samples. Samples were analysed as per the manufacturer’s instructions in a blinded manner without knowledge of diagnosis or other clinical information. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | Important baseline characteristics are adjusted for in statistical analyses. Factors adjusted for were age, gender, BMI, disease activity and ileum involvement. as potential confounders. However, it is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. Assessment of efficacy of tool is based on event rate of complication of disease or need for surgery. | Moderate |
| Clear definitions of the important confounders measured are provided | Yes |
| Measurement of all important confounders is adequately valid and reliable | Yes |
| The method and setting of confounding measurement are the same for all those in the study | Yes |
| Appropriate methods are used if imputation is used for missing confounder data | Not applicable |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) | No |
| Important potential confounders are accounted for in the analysis | Yes – see comment above |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | No model built. |
| The selected statistical model is adequate for the design of the study | Yes |
| There is no selective reporting of results | Yes |
| Abbreviations: IBD, inflammatory bowel disease; IBDX, Crohn’s disease Prognosis Test. | | | |

Table 6. Seow 2009

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Seow 2009 | | |
| Study details (journal, year, volume, page range) | *Am J Gastro*, 2009, 104, (6), 1426–1434. | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper. | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes  However, the study enrolled a mixed population of adults and children, those with a new diagnosis and an established diagnosis of CD, and varying degrees of existing complicated disease. Data for different subgroups are not reported separately. | Moderate |
| Description of the source population or population of interest | Yes |
| Description of the baseline study sample | Yes |
| Adequate description of the sampling frame and recruitment | Yes |
| Adequate description of the period and place of recruitment | Yes |
| Adequate description of inclusion and exclusion criteria | Yes |
| Study attrition | Adequate response rate for study participants | Yes  Number of people with a positive test for one or more antibody totals 378 out of 517 people for whom samples were available (73.1%). It is unclear from the details in the full publication whether the remaining 139 people did not test positive for an antibody, or whether their samples were not analysed. | Low |
| Description of attempts to collect information on people who dropped out | Not applicable |
| Reasons for loss to follow-up are provided | Not applicable |
| Adequate description of those lost to follow-up | Not applicable |
| There are no important differences between people who completed the study and those who did not | Not applicable |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes |
| Continuous variables are reported or appropriate cut points are used | Yes |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes |
| Appropriate methods of imputation are used for missing prognostic factor data | Not applicable |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Frozen serum samples were forwarded to Glycominds Limited for analysis in a masked manner. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | No  Confounders are not discussed in the publication.  It is unclear as to what extent antibody levels change over time in individual people and what factors influence changes in levels. | Moderate |
| Clear definitions of the important confounders measured are provided |
| Measurement of all important confounders is adequately valid and reliable |
| The method and setting of confounding measurement are the same for all those in the study |
| Appropriate methods are used if imputation is used for missing confounder data |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) |
| Important potential confounders are accounted for in the analysis |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | No model built. |
| The selected statistical model is adequate for the design of the study | Yes |
| There is no selective reporting of results | Yes |
| Abbreviation: IBD, inflammatory bowel disease; IBDX, Crohn’s disease Prognosis Test. | | | |

Table 7. Wolfel 2017

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Wolfel 2017 | | |
| Study details (journal, year, volume, page range) | *Gastroenterology*, 2017, 152, (5), S605 | | |
| Type of report (full paper/only abstract/conference abstract) | Conference abstract. | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | Yes.  All people have a diagnosis of CD and have undergone one CD-related surgery. | Unclear |
| Description of the source population or population of interest | No, insufficient detail provided in the abstract. |
| Description of the baseline study sample | No, only limited detail on age and gender provided in the abstract. |
| Adequate description of the sampling frame and recruitment | No, insufficient detail provided in the abstract. |
| Adequate description of the period and place of recruitment | No, insufficient detail provided in the abstract. |
| Adequate description of inclusion and exclusion criteria | No, insufficient detail provided in the abstract. |
| Study attrition | Adequate response rate for study participants | Unclear. | Unclear |
| Description of attempts to collect information on people who dropped out | Unclear. |
| Reasons for loss to follow-up are provided | Unclear. |
| Adequate description of those lost to follow-up | Unclear. |
| There are no important differences between people who completed the study and those who did not | Unclear. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes. | Unclear |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes. |
| Continuous variables are reported or appropriate cut points are used | Unclear. |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Unclear. |
| Adequate proportion of the study sample has complete data for the prognostic factor | Unclear. |
| Appropriate methods of imputation are used for missing prognostic factor data | Unclear. |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes. | Unclear |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Unclear. |
| The method and setting of outcome measurement are the same for all those in the study | Unclear. |
| Study confounding | Most important confounders are measured | The authors state that “Multivariable Cox regression analysis was performed to assess the associations between markers and recurrence of surgery adjusting for potential confounders”. Potential confounders are not listed. | Unclear |
| Clear definitions of the important confounders measured are provided |
| Measurement of all important confounders is adequately valid and reliable |
| The method and setting of confounding measurement are the same for all those in the study |
| Appropriate methods are used if imputation is used for missing confounder data |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) |
| Important potential confounders are accounted for in the analysis |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | No | Unclear |
| Strategy for model building is appropriate and is based on a conceptual framework or model | Unclear |
| The selected statistical model is adequate for the design of the study | Unclear |
| There is no selective reporting of results | Unclear |
| Abbreviation: IBD, inflammatory bowel disease. | | | |

#### PredictSURE-IBD

Table 8. Biasci 2019

|  |  |  |  |
| --- | --- | --- | --- |
| ***Reviewer and study information*** | | | |
| Reviewer name | Sam Barton | | |
| Study ID (Author name, year) | Biasci 2019 | | |
| Study details (journal, year, volume, page range) | *Gut*, 2019, (68), 1386–1395 | | |
| Type of report (full paper/only abstract/conference abstract) | Full paper | | |
| Domain | Aspects of trial for consideration in assessment of bias | Comment in support of assessment of bias | Rating of risk of bias |
| Study participation | Adequate participation in the study by eligible persons | People must have active disease for the tool to detect the desired sequence.  Validation cohort predominantly comprises those with newly diagnosed CD and disease is active. | Low |
| Description of the source population or population of interest | Yes. |
| Description of the baseline study sample | Not supplied for validation cohort in the full publication. Baseline characteristics provided by authors on request. |
| Adequate description of the sampling frame and recruitment | Yes. |
| Adequate description of the period and place of recruitment | Yes. |
| Adequate description of inclusion and exclusion criteria | Unclear reporting of inclusion criteria for validation cohort in full publication. Authors helpfully confirmed that inclusion criteria for validation cohort were the same as those for the training cohort. |
| Study attrition | Adequate response rate for study participants | Systematic literature search identified three conference abstracts that referred to a validation cohort comprising 85 people rather than the 66 reported in the full publication.146-148 During the DAP process, the company clarified that the cohort comprising 85 people refers to the validation work at an earlier stage of research and additional samples were added before publication of the full text. The EAG considers that it is unclear whether there are two cohorts or people have been lost to follow-up. | Unclear |
| Description of attempts to collect information on people who dropped out | Not reported. |
| Reasons for loss to follow-up are provided | Not reported. |
| Adequate description of those lost to follow-up | Not reported. |
| There are no important differences between people who completed the study and those who did not | Unclear. |
| Prognostic factor measurement | A clear definition or description of the prognostic factor is provided | Yes | Low |
| Method of prognostic factor measurement is adequately valid and reliable (i.e., direct ascertainment; secure record, hospital record) | Yes  Gene expression analyses. However, people must have active disease. |
| Continuous variables are reported or appropriate cut points are used | Not applicable |
| The method and setting of measurement of prognostic factor is the same for all those in the study | Yes |
| Adequate proportion of the study sample has complete data for the prognostic factor | Yes |
| Appropriate methods of imputation are used for missing prognostic factor data | Unclear |
| Outcome measurement | A clear definition of the outcome of interest is provided (including time of death) | Yes for the clinical outcome of time to event.  The authors of the systematic review appreciate that it will be difficult to determine the true clinical impact of the tool. | Low |
| Method of outcome measurement used is adequately valid and reliable (i.e., independent masked assessment, hospital record or record linkage) | Yes  Treating clinicians were masked to the biomarker results, and to gene expression analyses. |
| The method and setting of outcome measurement are the same for all those in the study | Yes |
| Study confounding | Most important confounders are measured | Assessment of efficacy of tool is based on time to an event involving treatment escalation based on clinical judgement. Confounders are not discussed in the full publication. | Unclear |
| Clear definitions of the important confounders measured are provided |
| Measurement of all important confounders is adequately valid and reliable |
| The method and setting of confounding measurement are the same for all those in the study |
| Appropriate methods are used if imputation is used for missing confounder data |
| Important potential confounders are accounted for in the study design (by limiting the study to specific population groups, or by matching) |
| Important potential confounders are accounted for in the analysis |
| Statistical analysis and reporting | Sufficient presentation of data to assess the adequacy of the analytic strategy | Yes | Low |
| Strategy for model building is appropriate and is based on a conceptual framework or model | In the development of the whole-blood sample test (IBDHi versus IBDLo), IBD1/IBD2 status was not included as a covariate in the batch normalisation of whole blood samples to reduce any downward bias in estimating the generalisation error during leave-one-out cross-validation. The impact of this is unclear. |
| The selected statistical model is adequate for the design of the study | Yes.  A statistical (machine) learning method was applied to the whole blood transcriptomic data to identify genes that could be used to calculate the probability of an individual belonging to the IBD1/IBD2 subgroups. |
| There is no selective reporting of results | Unclear. The full publication presents data for both the training and validation cohorts and the reporting of the data is considered to be unclear in some aspects. |
| Abbreviations: IBD, inflammatory bowel disease. | | | |

### Economic evaluations

Table 9. Drummond checklist for economic evaluations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Criteria | Study | | | | |
| Clark 2003 | Dretzke 2011 (TA187) | Hodgson 2018 NICE TA456 | Rafia 2016 NICE TA352 | NICE NG129 |
| 1. Was a well-defined question posed in answerable form? | Yes | Yes | Yes | Yes | Yes |
| 1.1. Did the study examine both costs and effects of the service(s) or programme(s)? | Yes | Yes | Yes | Yes | Yes |
| 1.2. Did the study involve a comparison of alternatives? | Yes (infliximab versus standard care) | Yes (adalimumab or infliximab versus standard care) | Yes (ustekinumab versus standard care) | Yes (vedolizumab versus standard care) | Yes induction of remission compared nine treatment strategies, and six treatments were compared for maintenance of remission) |
| 1.3. Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context? | Yes (NHS perspective) | Yes (NHS perspective) | Yes (NHS perspective) | Yes (NHS perspective) | Yes (NHS and personal social services perspective) |
| 2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)? | Yes | Yes | Yes | Yes | Yes |
| 2.1. Were there any important alternatives omitted? | No | No | No | No | No |
| 2.2. Was (should) a do-nothing alternative be considered? | No | No | No | No | No |
| 3. Was the effectiveness of the programme or services established? | Yes | Yes | Yes | Yes | Yes |
| 3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice? | Yes (ACCENT trial and Targan trial) | Yes (effectiveness for infliximab and adalimumab treatment were derived from ACCENT I and CHARM, respectively) | Yes (induction treatment assessed in UNITI-1, UNITI-2, CERTIFI and maintenance treatment assessed in IM-UNITI) | Yes (GEMINI II, GEMINI III) | Yes (various RCTs, treatment effectiveness of interventions for induction of and maintenance of remission derived from network meta-analyses) |
| 3.2. Was effectiveness established through an overview of clinical studies? | Yes | Yes | Yes | Yes | Yes |
| 3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results? | Yes (lack of observational data on the history of patients treated with infliximab has led to the reliance on data from one study, which involves two major assumptions: (i) QALY gains are reduced for people who revert to the more severe states, and (ii) the time patients spend in the various health states can be aggregated over their lifetimes, which, given the average age used, implies gains spread over about 40 years, which is a considerable extrapolation of the benefits of infliximab). | Yes (different assumptions in key studies made: Arsenau *et al*., Clark *et al*. and in the adalimumab model). | Yes (structural assumptions in the model inconsistent with UK clinical practice) | Yes (structural assumptions in the model influence outcomes) | No |
| 4. Were all the important and relevant costs and consequences for each alternative identified? | Yes | Yes | Yes | Yes | Yes |
| 4.1. Was the range wide enough for the research question at hand? | Yes | Yes | Yes | Yes | Yes |
| 4.2. Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints may also be relevant depending upon the particular analysis.) | Yes | Yes | Unclear | Yes | Yes |
| 4.3. Were the capital costs, as well as operating costs, included? | Unclear | Yes | Yes | Unclear | Yes |
| 5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life years)? | Yes | Yes | Yes | Yes | Yes |
| 5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis? | No | Unclear | No | No | No |
| 5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately? | Unclear | No | Unclear | No | No |
| 6. Were the cost and consequences valued credibly? | Yes | Yes | Yes | Yes | Yes |
| 6.1. Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers’ views and health professionals’ judgements) | Yes | Yes | Yes | Yes | Yes |
| 6.2. Were market values employed for changes involving resources gained or depleted? | Unclear | Unclear | Unclear | No | Unclear |
| 6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values? | Unclear | Unclear | Yes | Unclear | Unclear |
| 6.4. Was the valuation of consequences appropriate for the question posed (i.e. has the appropriate type or types of analysis – cost-effectiveness, cost-benefit, cost-utility – been selected)? | Yes | Yes | Yes | Yes | Yes |
| 7. Were costs and consequences adjusted for differential timing? | Unclear | Unclear- UK study from 2004 calculated the costs of CD. Unknown if adjusted. | Unclear | Unclear | Unclear |
| 7.1. Were costs and consequences that occur in the future ‘discounted’ to their present values? | Yes (6 % for costs and 1.5% benefits respectively) | No | Yes (3.5% for costs and benefits) | Yes (3.5% for costs and benefits) | Yes (3.5% for costs and benefits) |
| 7.2. Was there any justification given for the discount rate used? | No | No discounting required for 1-year time horizon. | Yes (in accordance with NICE reference case) | Yes (in accordance with NICE guidance) | Yes (in accordance with NICE guidance) |
| 8. Was an incremental analysis of costs and consequences of alternatives performed? | Yes | Yes | Yes | Yes | Yes |
| 8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated? | Yes | Yes | Yes | Yes | Yes |
| 9. Was allowance made for uncertainty in the estimates of costs and consequences? | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) |
| 9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed? | NA | NA | NA | NA | NA |
| 9.2. If a sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)? | Yes | Yes | Yes | Yes | Yes |
| 9.3. Were the study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)? | Yes (chronic active model was highly sensitive to rate of ‘flare’ for episodic treatment. The flare rate chosen was 10%, which seemed reasonable based on clinical opinion. If more frequent flare was seen, then costs increased substantially) | Yes | Yes (model was sensitive to the duration of treatment and the analytic time horizon) | Yes | Yes |
| 10. Did the presentation and discussion of study results include all issues of concern to users? | Yes | Yes | Yes | Yes | Yes |
| 10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion? | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) |
| 10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology? | Yes | Yes | Yes | No | No |
| 10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups? | No | Yes | Yes | Yes | Yes |
| 10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)? | No | Yes | Yes | Yes | Yes |
| 10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the ‘preferred’ programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes? | No | No | No | No | No |

Table 10. Drummond checklist for economic evaluations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Criteria | Study | | | | | |
| Marchetti 2013 | Freeman 2016 | Saito 2013 | Bodger 2009 | Loftus 2009 | Lindsay 2008 |
| 1. Was a well-defined question posed in answerable form? | Yes | Yes | Yes | Yes | Yes | Yes |
| 1.1. Did the study examine both costs and effects of the service(s) or programme(s)? | Yes | Yes | Yes | Yes | Yes | Yes |
| 1.2. Did the study involve a comparison of alternatives? | Yes (top-down versus step up) | Yes (monitoring of serum anti-TNF-alpha antibody levels versus no testing/standard care) | Yes (infliximab monotherapy versus infliximab plus azathioprine) | Yes (infliximab and adalimumab for versus standard care) | Yes (adalimumab versus non biologic therapies in maintenance of CD) | Yes (infliximab versus standard care) |
| 1.3. Was a viewpoint for the analysis stated and was the study placed in any particular decision-making context? | Yes (Italian Healthcare System) | Yes (NHS perspective) | Yes (NHS perspective) | Yes (NHS perspective) | Yes (NHS perspective and from the perspective of the social decision maker) | Yes (NHS perspective) |
| 2. Was a comprehensive description of the competing alternatives given (i.e. can you tell who did what to whom, where, and how often)? | Yes | Yes | Yes | Yes | Yes | Yes |
| 2.1. Were there any important alternatives omitted? | No | No | No | No | No | No |
| 2.2. Was (should) a do-nothing alternative be considered? | No | No | No | No | No | No |
| 3. Was the effectiveness of the programme or services established? | Yes | Yes | Yes | Yes | Yes | Yes |
| 3.1. Was this done through a randomised, controlled clinical trial? If so, did the trial protocol reflect what would happen in regular practice? | Yes (trial by D'Haen's *et al*.) | Yes (ACCENT I, ACCENT II) | Yes (ACCENT I, SONIC, Lemann trial) | Yes (CHARM and ACCENT I) | Yes (CHARM and CLASSIC 1) | Yes (ACCENT I, ACCENT II, Targan trial, Present trial) |
| 3.2. Was effectiveness established through an overview of clinical studies? | Yes | Yes | Yes | Yes | Yes | Yes |
| 3.3. Were observational data or assumptions used to establish effectiveness? If so, what are the potential biases in results? | Unclear | Yes | Yes | Yes (surgical rates based on observational data) | Yes | Yes |
| 4. Were all the important and relevant costs and consequences for each alternative identified? | Yes | Yes | Yes | Yes | Yes | Yes |
| 4.1. Was the range wide enough for the research question at hand? | Yes | Yes | Yes | Yes | Yes | Yes |
| 4.2. Did it cover all relevant viewpoints? (Possible viewpoints include the community or social viewpoint, and those of patients and third-party payers. Other viewpoints may also be relevant depending upon the particular analysis.) | Yes | Yes | Yes | Yes | Yes | Yes |
| 4.3. Were the capital costs, as well as operating costs, included? | Yes | Yes | Yes | Yes | Yes | Yes |
| 5. Were costs and consequences measured accurately in appropriate physical units (e.g. hours of nursing time, number of physician visits, lost work-days, gained life years)? | Yes | Unclear | Yes | Yes | Yes | Yes |
| 5.1. Were any of the identified items omitted from measurement? If so, does this mean that they carried no weight in the subsequent analysis? | Unclear | Unclear | Unclear | Unclear | No | No |
| 5.2. Were there any special circumstances (e.g., joint use of resources) that made measurement difficult? Were these circumstances handled appropriately? | No | No | Unclear | No | No | Yes (true placebo effect could not be estimated from the ACCENT trials) |
| 6. Were the cost and consequences valued credibly? | Yes | Yes | Yes | Yes | Yes | Yes |
| 6.1. Were the sources of all values clearly identified? (Possible sources include market values, patient or client preferences and views, policy-makers’ views and health professionals’ judgements) | Yes | Yes | Yes | Yes | Yes | Yes |
| 6.2. Were market values employed for changes involving resources gained or depleted? | Unclear | Unclear | Unclear | Unclear | Unclear | Unclear |
| 6.3. Where market values were absent (e.g. volunteer labour), or market values did not reflect actual values (such as clinic space donated at a reduced rate), were adjustments made to approximate market values? | Unclear | Unclear | Unclear | Unclear | No | Yes |
| 6.4. Was the valuation of consequences appropriate for the question posed (i.e. has the appropriate type or types of analysis – cost-effectiveness, cost-benefit, cost-utility – been selected)? | Yes | Yes | Yes | Yes | Yes | Yes |
| 7. Were costs and consequences adjusted for differential timing? | Unclear | Yes | Yes | Yes | Yes | Yes |
| 7.1. Were costs and consequences that occur in the future ‘discounted’ to their present values? | Yes (3.5% for costs and benefits) | Yes (3.5% for costs and benefits) | NA | Yes (3.5% for costs and benefits) | Yes (3.5% for costs and benefits) | Yes (3.5% for costs and benefits) |
| 7.2. Was there any justification given for the discount rate used? | Yes (international guidelines) | Yes (NICE reference case) | NA | Yes (NICE reference case) | No | Yes (NICE reference case) |
| 8. Was an incremental analysis of costs and consequences of alternatives performed? | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) |
| 8.1. Were the additional (incremental) costs generated by one alternative over another compared to the additional effects, benefits, or utilities generated? | Yes | Yes | Yes | Yes | Yes | Yes |
| 9. Was allowance made for uncertainty in the estimates of costs and consequences? | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) | Yes (sensitivity analysis) |
| 9.1. If data on costs and consequences were stochastic (randomly determined sequence of observations), were appropriate statistical analyses performed? | NA | NA | NA | NA | NA | NA |
| 9.2. If a sensitivity analysis was employed, was justification provided for the range of values (or for key study parameters)? | Yes | Yes | Yes | Yes | Yes | No |
| 9.3. Were the study results sensitive to changes in the values (within the assumed range for sensitivity analysis, or within the confidence interval around the ratio of costs to consequences)? | Yes | Yes (sensitive to a 10 percent increase in the utility value for patients who regain response in both reflex and concurrent testing) | Yes (analyses showed that the quality of life utility associated with nonresponding active disease was the most influential parameter on the cost-effectiveness of the therapies) | Yes | Yes | Yes (in OWSA, because of the weight-based dosing of infliximab, patient weight had the most impact on the ICER) |
| 10. Did the presentation and discussion of study results include all issues of concern to users? | Yes | Yes | Yes | Yes | Yes | Yes |
| 10.1. Were the conclusions of the analysis based on some overall index or ratio of costs to consequences (e.g. cost-effectiveness ratio)? If so, was the index interpreted intelligently or in a mechanistic fashion? | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) | Yes (cost per QALY gained) |
| 10.2. Were the results compared with those of others who have investigated the same question? If so, were allowances made for potential differences in study methodology? | No | Yes | Yes | Yes | Yes | Yes |
| 10.3. Did the study discuss the generalisability of the results to other settings and patient/client groups? | Yes | Yes | Yes | Yes | Yes | Yes |
| 10.4. Did the study allude to, or take account of, other important factors in the choice or decision under consideration (e.g. distribution of costs and consequences, or relevant ethical issues)? | No | Yes | Yes- | Yes | No | Yes |
| 10.5. Did the study discuss issues of implementation, such as the feasibility of adopting the ‘preferred’ programme given existing financial or other constraints, and whether any freed resources could be redeployed to other worthwhile programmes? | No | No | No | No | No | No |