**Title: An explanation of the documents contained in the Short Course Oncology Treatment (SCOT) trial deposit. .**

Institution: CRUK Clinical Trials Unit, Glasgow, United Kingdom

Project manager: Andrea Harkin

Computer manager: Geraldine McBride

Study Statistician: Jim Paul

**Description**

The patient level data that was collected for the purposes of the SCOT trial is held by the CRUK Clinical Trials Unit (CTU) in Glasgow.

This deposit contains:

1. Examples of the templates that were used to collect this data and
2. A separate list of the data variables that have been collected.

There is no patient level data held in this repository. If you would like to access this patient level data, please submit a formal request to the Glasgow CRUK CTU using the contact details provided. There will be occasions where patient level data contains identifiable information. In these cases, the data corresponding to these specific, patient-identifying data variables will not be shared with other parties.

1. **Templates for data collection**

There are 11 templates for data collection which are available to view in this deposit.

**Template 1: Randomisation Form v4\_00712.**

Collects information at randomisation.

**Template 2: Wk 12 Randomisation Form v2 130608**

Collects information, including screening criteria details, for those patients randomised at week 12 of treatment.

**Template 3: Treatment Form\_v4\_240513**

Collects information on cycle number, type of treatment, side effects, resource use, if QOL questionnaires have been completed and if adverse events are associated with treatment withdrawal.

**Template 4: Preg Notification Form v1 080208**

Collects information if a patient on the trial becomes pregnant.

**Template 5: High risk stage II Form\_V1.0\_29Jan2013**

Collects extra information for those patients who have high risk stage II colorectal cancer (versus stage III).

**Template 6: EORTC QLQ-C30 v2 060309**

Collects information on the timing and results for the responses to the EORTC QLQ-C30 quality of life questionnaire.

**Template 7: EQ-5D Questionnaire v2 150409**

Collects information on the timing and results for the responses to the EQ5D quality of life questionnaire.

**Template 8: GOG Ntx 4\_V3\_030513**

Collects information on the timing and results for the responses to the neurotoxicity questionnaire.

**Template 9: Follow up Form\_V3\_240513**

Collects information from each follow up visit (including imaging done, resource use, QOL questionnaire completion and disease recurrence status).

**Template 10: Long\_term\_Follow Up Form V2Nov2018**

Collects data for those patients on long term follow up including disease recurrence and survival status.

**Template 11: Consent withdrawal notification form V2 010408**

Collects details for any patient who withdraws consent to participate and have their information collected as part of the SCOT trial.

**List of data variables collected** There are 25 separate lists of data variables which are available to view in this deposit in the Master Document. Each list contained in the Master Document has a title in the format: Trial Code (GI145)\_Title of data collection\_Version.

The raw data for each of the data variables listed will generally be collected in one the following formats:

* The study number (GI145);
* A number value – the number value format will be recorded as follows,

Example: (Number f99.9), where f= fixed format, 99 is the number of digits expected before the decimal point (in this case 2), a decimal point, and 9 is the number of digits expected after the decimal point (in this case 1);

* Free text;
* A date;
* The answer options in relation to the question for e.g. 1;2;3;
* The answer option in relation to a question which can have a tick for the answer will be coded as 1=Ticked; Null=Not ticked.

The number of data variable fields included in the patient level dataset (excluding study number, trial number, form type and form sequence) and a short description of the groups of data variables available is outlined below. Below is a summary of the Master Document which contains 25 lists which each provide a detailed description of the data variable names the coding used for each variable when it is collected at the patient level.

 **File 1: GI145\_SITRAN\_V6**

These data variables include details of the recruiting site, CTU and time-point for randomisation for each patient.

Number of field names: 10

 **File 2: GI145\_REG\_V6**

These data variables outline baseline registration details for each patient. This includes the dates and results for clinical screening investigations, consent taken, an indication if baseline questionnaires have been completed and the patient trial number.

Number of field names: 57.

 **File 3: GI145\_MASTER\_REG\_V6**

This MASTER\_REG document is a record of the registration details for each patient that the CTU originally receives from each recruiting site. If the recruiting site subsequently identify errors relating the information provided, this master copy is kept as a permanent record of the original information received but the record will be updated with the correct information in File 2. File 2 is therefore the working file and File 3 will not be made available users of the data repository. **File 4: GI145\_W12REG\_V6**

These data variables include registration details for those patients who were randomised at week 12 rather than prior to week 1 of treatment. The delayed randomisation question period was from March 2008 to July 2009. During this period, some patients were randomised upfront (ie. Before start of treatment) and other patients were randomised at week 12 of treatment. This was determined by the recruiting site. We randomly selected sites to either randomise upfront or at week 12. After July 2009, all patients were randomised upfront.

Number of field names: 31

 **File 5: GI145\_CNWI\_5**

These data variables outline details for of the timing and reasons for withdrawal of consent by patients in the trial.

Number of field names: 15

 **File 6: GI145\_QOL\_V6**

These data variables include the timing of when patients completed quality of life questionnaires and includes the responses for each patient.

Number of field names: 8

 **File 7: GI145\_EQ5D\_V6**

These data variables include the timing and results for responses to the EQ-5D questionnaire.

Number of field names: 8

 **File 8: GI145\_NEUR\_V6**

These data variables include the timings and the results for responses to the GOG NTX 4 questionnaire.

Number of field names: 3

 **File 9: GI145\_CHEMO\_V6**

These data variables show the chemotherapy delivered to each patient. The time-point for chemotherapy delivered is divided into the first 6 weeks, weeks 7-12, weeks 13-18 and weeks 19-24. There are details on the patients’ height, weight, if chemotherapy was delayed and if so, for what reason, the names and doses of chemotherapy delivered, if any significant adverse events occurred and if the patient received intravenous calcium or magnesium supplements during chemotherapy administration.

Number of field names: 56

 **File 10: GI145\_TRTCHNG\_V6**

This data variables include details on treatment changes for each patient. The dataset indicates if a patient has discontinued oxaliplatin chemotherapy or if they have switched from capecitabine to intravenous 5-fluorouracil during the trial and if so, the reasons for doing so.

Number of field names: 9

 **File 11: GI145\_AE**

These data variables include details of the timing and severity of adverse reactions and other toxicities that have occurred during the trial for each patient.

Number of field names: 7

 **File 12: GI145\_HEA\_V6**

These data variables include details on resource use and hospital admissions for use in health economic analyses.

Number of field names: 7

 **File 13: GI145\_CHQDT\_V6**

These data variables indicate if each patient had completed quality of life questionnaires at specific time-points during the trial treatment phase and if not, the reasons why the form has not been completed.

Number of field names: 21

 **File 14: GI145\_FUQDT\_V6**

These data variables indicate if each patient had completed quality of life questionnaires at specific time-points during the trial follow up phase and if not, the reasons why the form has not been completed.

Number of field names: 6

 **File 15: GI145\_FCH\_V6**

These variables indicate, for specific time-points, if each patient was continuing trial treatment, finished trial treatment, discontinued treatment permanently before completing treatment or if they were due to be randomised at 12 weeks, but this did not occur.

Number of field names: 4

 **File 16: GI145\_WITH\_V6**

These data variables indicate the reason for early withdrawal of treatment for each patient who stopped trial treatment before completing the allocated treatment.

Number of field names: 23

 **File 17: GI145\_MIX\_V6**

These data variables include a mixture of variables for each patient including an indication of if they had anti-cancer treatment in the time between the current date and the last trial assessment, if treatment follow up has been stopped and if the patient has a confirmed previous recurrence.

Number of field names: 6

 **File 18: GI145\_SURV\_V6**

These data variables includes survival status for each patient and cause of death where relevant.

Number of field names: 8

 **File 19: GI145\_SLV\_V6**

These data variables include details on treatments that each patient may have received after trial treatment and the dates during which these treatments were administered.

Number of field names: 10

 **File 20: GI145\_REC\_V6**

These data variables contains information on imaging studies performed during follow up and details of disease recurrence for patients where relevant.

Number of field names: 18

 **File 21: GI145\_SGN\_V6**

These data variables indicate if there is a signature on the consent withdrawal form for patients with high risk stage II disease.

Number of field names: 2

 **File 22: GI145\_COM\_V6**

These data variables are a free text dataset for any further communication on each patient that has not been captured in the other forms.

Number of field names: 1

 **File 23: GI145\_PWQ\_V6**

These data variables relate to the responses to a questionnaire that is completed when a patient withdraws from the study (PWQ: Patient Withdrawal Questionnaire).

Number of field names: 4

 **File 24: GI145\_HR\_V6**

These data variables indicate the reasons why a patient has been deemed to have “high risk” stage II colorectal cancer and if they have had radiotherapy pre-or post-operatively as part of their first line treatment.

Number of field names: 14

 **File 25: GI145\_Long\_Term\_FUP\_V6**

These data variables include details on recurrence, survival status and cause of death where relevant, for those patients on long term follow up.

Number of field names: 13