**To:** [Trial PI]

**CC:** [Trial Statistician, Trial Data Manager, Magdalena Thurin, CIDC, CIMACs, FNIH if applicable]

**From:** CIMAC-CIDC Network, NCI

**Date:** [DATE]

**Re: Clinical Data Request for [Protocol #]** *“[Protocol title]”*

The CIMAC-CIDC Network requests transfer of the clinical data described in this letter, and in detail in **Appendix A**, for use in correlative studies.

When ready, please post these clinical data from trial **[Protocol #]** to the following secure webpage in NIH Box: [link to be provided when this data request is finalized]

Please notify [CIDC email address] when the data are posted.

**Delivery timing:**

Please compile, QC, and transfer de-identified clinical data for all patients in the trial, including those who were eligible but did not receive study treatment.

The **preferred version** of the clinical data is the data from the cleaned, locked dataset to be used in the primary publication for the trial – i.e., the clinical data to be used in the publication that reports the primary clinical outcomes from the trial, which could include secondary endpoints as well.

**If additional clinical data** are used in the correlative analysis, i.e., in addition to those elements listed in Appendix A, please let us know so they can also be posted.

If you find additional data column(s) need to be added for a data sheet to make sense or to further inform the analysis, please add the column(s) and explain them in the Data Dictionary.

The data can be delivered in two sequential phases if needed – Delivery A and Delivery B – as indicated in Appendix A, although one Delivery is preferred, to the extent possible:

* **Delivery “A”:** Clinical data to be delivered as soon as feasible (should not await Delivery B).
* **Delivery “B”:** Data to be sent in a later update of the data. Please provide updated files for all elements in Appendix A when providing Delivery B, to ensure the same lock date across all data.
* **Delivery “NA”:** The clinical data are not applicable to the trial or will likely never be available.

**Critical considerations:**

* **Data must be de-identified, including IDs**. No PHI may be accepted by the CIMAC-CIDC.
* Please provide **separate files for each data category** noted in Table 1 below (“**Category**” column). Files should contain data for those data elements as outlined in Appendix A.
* Please verify all patient identifiers can be linked to specimen data sent by the biorepository – i.e., **the Participant ID in the clinical data must match the Participant ID used in the specimen manifests**. This ID should be **blinded** – i.e., it should not be the operational ID actually used during the trial.
* **Please include the Participant ID in each row of all files submitted.**
* Please **label all data columns** with the labels contained in Appendix A’s “**Data Label**” column.If data differ from what is requested, please follow the instructions at the end of this document for assigning a Data Label to accommodate trial-specific data.
* You can include **multiple rows** per patient as needed in the data files, except in those instances where adding columns would be preferable (example provided at end of document).
* Please provide data in .csv or .xlsx format. **.csv format is strongly preferred.**
* A **clinical data dictionary**, as described in Appendix A, must be provided (as a separate document) with the clinical data, to describe **all variables** provided in the data. To the extent possible, the order of the variables in the Data Dictionary should mirror the order of the variables in the data files. The data dictionary should be a separate document.
  + An example Data Dictionary is provided in Appendix A, in the third tab. It includes **Data Label** (e.g., A1), **Variable** (variable name used in trial database), **Variable label** (brief variable description, such as Data Element name in Appendix A’s Column C or similar, as appropriate to the trial.), **Code**, **Code definition** (e.g., possible values), and how a derived variable was **derived or calculated**. NOTE: Variable names should not contain spaces. Underscores can be used to fill spaces if needed (for example, “Treatment arm” would be “Treatment\_arm”).
  + **Blank data fields** should be defined in the Data dictionary. A one-time explanation in the Data Dictionary is appropriate if all blanks have a consistent meaning (e.g., missing data).
* **Dates should not be provided.** All time-related events must be relative to an **anchor date** (i.e., the start date from which the time-related event was calculated). We recommend using the **date of registration** (study enrollment) as the anchor for most categories. Please use plus or minus ***days*** relative to the anchor. Please note the anchor date for each time-related element in the **Data Dictionary**.

After you post the data, CIDC will QC it and may reach out to you with questions from their QC.

**Table 1. Data Categories Requested**

***Further details provided in Appendix A, attached.***

|  |  |  |
| --- | --- | --- |
| **Data Level** | **Category (submitted in separate files)** | **Data Label/ Column Headers** |
| Study-level | Data Dictionary |  |
|  | CRFs |  |
|  | Study [Study-Level Descriptors] | S |
| Patient-level | Demographics | A |
|  | History | B |
|  | History\_Description | B |
|  | History\_GVHD [for heme trials] | B |
|  | Comorbidities | As assigned by study team |
|  | Molecular\_Features | As assigned by study team |
|  | Disease | C |
|  | Response | D |
|  | Treatment | E |
|  | Additional\_Treatment | E |
|  | Treatment\_Dose | K |
|  | Specimen\_Collection | M |
|  | Adverse\_Event | F |
|  | GVHD\_Post-enrollment [for heme trials] | F |
|  | Specialized\_Data | G |

**Goal of the Appendix A format**

* Appendix A is not intended as a mapping requirement for study teams. Rather, it is structured to guide submission of native data in a format that facilitates harmonization of data across Network trials.
* PLEASE NOTE: Once Appendix A is annotated by the trial team, including with any additional data elements, CIDC will use Appendix A as an “inventory” to help them track which data have been submitted for the CIMAC-CIDC studies, and which have yet to be submitted.
* Data Labels have been standardized across Common Data Elements (CDEs). Each data Category (e.g., Demographics) has been designated a letter in its Data Label, as shown in Table 1 above. Demographic data, for example, are comprised of CDEs A1 – A14.
* Appendix A includes reference to CDEs from the NCI Cancer Data Standards Registry and Repository (caDSR; CDE Browser: <https://cdebrowser.nci.nih.gov>). Adhering to caDSR data elements, to the extent possible, will facilitate sharing of CIMAC data sets through NCI.

**Expanding/modifying the Data Label (column header) to accommodate trial-specific data**

1. For some variables with more than one piece of data, adding columns instead of rows might be preferable. In these cases, you can add an **underscore and number (\_#)** to the Data Label in the **column header**, as illustrated in the example below for a patient reporting multiple races.

**Example:**

|  |  |  |  |
| --- | --- | --- | --- |
| **A1** | **A2\_1** | **A2\_2** | **A2\_3** |
| (Patient identifier) | Black or African American | American Indian or Alaska Native | Asian |

**NOTE:** For other data files, however, **multiple rows per patient** may be preferable for a given data element, instead of multiple columns. For example, multiple rows per patient would be preferred in the files for Additional\_Treatment, Specimen\_Collection, History\_Description, Comorbidities, Treatment\_Dose, Adverse\_Event, and GVHD\_Post-enrollment.

1. If a **trial-specific definition of a concept differs** from what is requested, add an **underscore and “Alt” (“\_Alt”)** to the Data Label/Column header, as described in the example below. Additionally, please define the data element in the Data Dictionary.

**Example:**

The trial’s overall “Study Outcome, Objective response rate” is requested in CDE **S1** (if known). It is defined as the proportion of patients with Complete Response or Partial Response per RECIST. A trial, however, uses 6-month Progression-Free Survival as its primary endpoint, not Objective response rate. This trial would use the Data Label (column header) **S1\_Alt** instead of S1 and define the 6-month Progression-Free Survival measure in the Data Dictionary.

1. For trial-specific data submitted that are **outside the common data set listed in Appendix A**, please use the data labels from the native trial dataset,i.e., the trial data labels. Please also use trial data labels for data under “Comorbidities”, “Molecular\_Features”, and “Non-CIMAC Molecular Tests During Study”.