MOHCCN Clinical Data Model v2 Guidelines

Table of Contents

1. Data Dictionary

- Field Descriptors
- Permissible Values
- Tumour Staging Classifications
- Tumour Grading Classifications
- Response to Treatment Criteria
- Submitting multiple values: Certain fields accept multiple values (these are indicated in the "Notes" column). Multiple values can be submitted using a pipe character ("|") as a delimiter.

2. Clinical Data Validation Rules

- Identifier fields
- Primary Diagnosis, Treatment and Follow Up Identifiers
- Submitting Missing Values for Extended Clinical Fields
- Cross Field Validations

3. Basic Definitions

Donor: A living or deceased individual who is the source of the specimen in accordance with established medical criteria, procedures and privacy regulations (Reference ISBER)

Specimen: A specimen is a specific tissue, blood, urine, or other material collected for analysis or a small fragment of tissue for microscopic study, taken from a single donor at a specific time. (Reference ISBER)

Sample: A single unit containing material derived from one specimen. (Reference ISBER)

4. Sample Registration

Sample registration is the first step in the data submission life cycle and requires the data submitter to submit a basic set of data for each sample before submitting any clinical or molecular data. This is to ensure that relationships between different entities are maintained across all molecular and clinical data submissions.
During sample registration, each Donor, Specimen and Sample entity will be assigned an ID that maps to your program's internal identifier (also referred to as a "submitter_id"). Any attempts to submit molecular or clinical data that does not refer to a registered donor, specimen or sample will result in an error.

Example: Submission of tumour/normal pair of samples for submitter_donor_id "donor1"

program_id	submitter_donor_id	submitter_specimen_id	specimen_tissue_source	tumour_normal_designation	specimen_type	submitter_sample_id	sample_type
BRST-CA	donor1	specimen 1a	Solid tissue	Tumour	Primary tumour	sample1 T	Total DNA
BRST-CA	donor1	specimen 1b	Solid tissue	Normal	Normal	sample1 N	Total DNA

- Fields used for data processing purposes:
    - "tumour_normal_designation": to identify tumour/normal pair for a given donor
    - "sample_type": the value in this field determines which pipeline the molecular data will be processed in (ie. DNA vs. RNA pipeline).

5. Donor

The collection of data elements related to a specific donor in a program.

How to indicate a donor is lost to follow up

If a donor was lost to follow up, the data submitter can indicate after which clinical event the donor became lost to follow up in the "lost_to_followup_after_clinical_event_identifier" field. Since each clinical event is associated with a date, we will know when the donor was last known

to be alive. The "is_deceased" must be "No" if the donor is lost to follow up (ie. if donor is known to be deceased, then they can't be lost to follow up)

Example 1:

Donor lost to follow up after their last treatment

Lost to follow up after treatment

Donor was lost to follow up after surgery treatment (TR_01)
Donor was last known to be alive 2017-10

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	TR_01

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10

Follow Up

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup

Donor was lost to follow up after their last treatment (TR_01). This means the donor was last known to be alive 2017-10.

Examples of how to update submission if a donor that was indicated as lost to follow up was found again later.

Example 1: Donor is found to be alive after they had been indicated as lost to follow-up:

In this example, the data submitter must update the "lost_to_followup_after_clinical_event_identifier" field to be empty.

Example submission:

Donor lost to follow up after their last follow up

Lost to follow up after last follow up

Donor was lost to follow up after their last follow up (FL22)
Donor was last known to be alive 2018-04

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	FL22

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10
DN_10	TR_02	PD_01	Chemotherapy	2017-11	2018-04

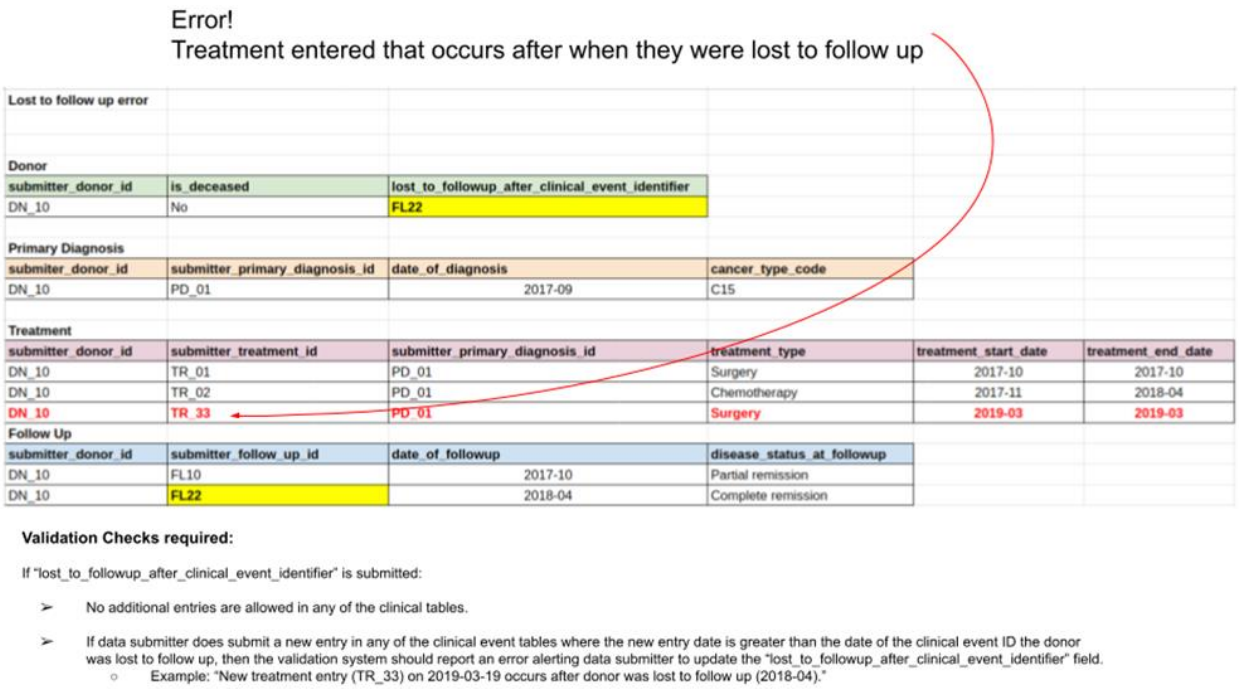
Follow Up

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup
DN_10	FL10	2017-10	Partial remission
DN_10	FL22	2018-04	Complete remission

Donor was lost to follow up after their last follow up (“FL22”). This means the donor was last known to be alive 2018-04.

Donor was lost to follow up after their last follow up appointment (“FL22”) in April 2018 and data submitter indicated this by submitting “lost_to_followup_after_clinical_event_identifier” = “FL22”.

The donor was found again a year later when they had surgery in March 2019. The data submitter submits the surgery treatment as a new row in the Treatment table (submitter_treatment_id = “TR33”), but since they did not update the “lost_to_followup_after_clinical_event_identifier” field (ie. update to be empty), the clinical validation will fail.



Lost to follow up error

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	FL22

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10
DN_10	TR_02	PD_01	Chemotherapy	2017-11	2018-04
DN_10	TR_33	PD_01	Surgery	2019-03	2019-03

Follow Up

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup
DN_10	FL10	2017-10	Partial remission
DN_10	FL22	2018-04	Complete remission

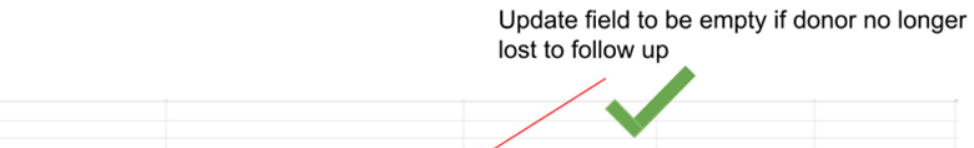
Validation Checks required:

If "lost_to_followup_after_clinical_event_identifier" is submitted:

- No additional entries are allowed in any of the clinical tables.
- If data submitter does submit a new entry in any of the clinical event tables where the new entry date is greater than the date of the clinical event ID the donor was lost to follow up, then the validation system should report an error alerting data submitter to update the "lost_to_followup_after_clinical_event_identifier" field.
  - Example: "New treatment entry (TR_33) on 2019-03-19 occurs after donor was lost to follow up (2018-04)."

To correct the submission, the data submitter will need to update the "lost_to_followup_after_clinical_event_identifier" field to be empty, since the donor is no longer lost to follow up.

To correct:



Lost to follow up error

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

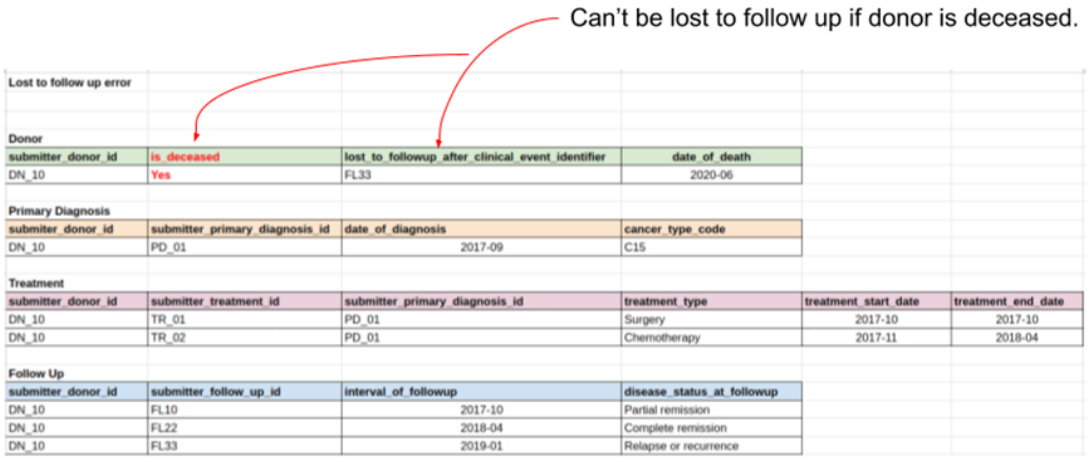
submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10
DN_10	TR_02	PD_01	Chemotherapy	2017-11	2018-04
DN_10	TR_33	PD_01	Surgery	2019-03	2019-03

Follow Up

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup
DN_10	FL10	2017-10	Partial remission
DN_10	FL22	2018-04	Complete remission

Example 2: Donor was found to have died 6 months after they were lost to follow up.

If a donor is discovered to be deceased after they were lost to follow up ("is_deceased" = "Yes"), then the "lost_to_followup_after_clinical_event_identifier" field should not be submitted.



Lost to follow up error

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier	date_of_death
DN_10	Yes	FL33	2020-06

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10
DN_10	TR_02	PD_01	Chemotherapy	2017-11	2018-04

Follow Up

submitter_donor_id	submitter_follow_up_id	interval_of_followup	disease_status_at_followup
DN_10	FL10	2017-10	Partial remission
DN_10	FL22	2018-04	Complete remission
DN_10	FL33	2019-01	Relapse or recurrence

Validation Checks Required:

"lost_to_followup_after_clinical_event_identifier":

- Cannot be submitted if "is_deceased" = "Yes"
- Can only be submitted if "is_deceased" is not "Yes"

If data submitter discovers the donor is deceased after they had been lost to follow up, the "lost_to_followup_after_clinical_event_identifier" field must be updated to be empty since, and the date of death should be submitted in the "date_of_death" field.

How to denote cancer-specific survival vs. overall survival (patient is known to be alive, but don't know the status of cancer).

Example use case: Donor had a follow up appointment at the cancer clinic in August 2015 where their disease status was found to be complete remission. The donor became lost to follow up after that August 2015 follow up appointment. Two years later, the cancer clinic found out that the donor showed up for an ophthalmology appointment somewhere in January 2017, but they don't have any information about the donor's cancer disease status.

1. Data submitter must indicate when the donor became lost to follow up in their cancer-related timeline. In this example, the donor became lost to follow up after their last appointment at the cancer clinic in August 2015 which is submitter_follow_up_id = "SF0012".

Donor

submitter_donor_id	vital_status	lost_to_followup_after_clinical_event	date_alive_after_lost_to_followup
Donor123	Alive	SF0012	

Follow Up

submitter_follow_up_id	submitter_donor_id	date_of_followup	disease_status_at_follow up
SF0012	Donor123	2015-08	Complete remission

2. Clinic discovers the donor showed up at an ophthalmology appointment 2 years later in January 2017, but they don't have any cancer disease status information on the donor, they only know the donor is still alive as of January 2017.

Donor

submitter_donor_id	vital_status	lost_to_followup_after_clinical_event	date_alive_after_lost_to_followup
Donor123	Alive	SF0012	2017-01

Follow Up

submitter_follow_up_id	submitter_donor_id	date_of_followup	disease_status_at_followup
SF0012	Donor123	2015-08	Complete remission

6. Specimen

- Both tumour and normal specimens must be submitted in this table. Tumour-specific fields such as "tumour_histological_type" should not be submitted if the specimen is normal. Clinical validation will fail if tumour-specific fields are submitted for normal specimens, or if tumour-specific fields are missing for tumour specimens.
- "submitter_donor_id" and "submitter_specimen_id" will be validated to ensure the relationship is consistent.
- Either pathological (in Specimen table) or clinical staging (in Primary Diagnosis table) must be submitted. If both pathological staging and clinical staging is available, then please submit both as opposed to just one staging.
- The "pathological_stage_group" will be cross-validated against "pathological_tumour_staging_system". Refer to Tumour Staging Classifications documentation to ensure the correct stage group term is submitted.
- If pathological staging is associated with primary diagnosis, then

"specimen_collection_date" should be the same date as "date_of_diagnosis".

- The "tumour_grade" field will be cross-validated against the "tumour_grading_system" field. Refer to the Tumour Grading Classifications documentation to ensure the correct tumour grade term is submitted.

7. Primary Diagnosis

- "date_of_diagnosis": Submit the date of the *earliest* diagnosis of cancer. For example, a donor can be given neoadjuvant treatment before the definitive diagnosis, so the date of earliest diagnosis should be submitted.

- "basis_of_diagnosis": The suggested codes from IARC Recommendations on registry practices are hierarchical, so that the higher number represents the more valid basis, and should thus be used for this purpose. If there is no information on how the diagnosis had been made (information obtained from an automated source, for example) "Unknown" should be submitted.

Table 1. Basis of diagnosis codes

Code	Description	Criteria
0	**Death certificate only**	The only information to the registry is from a death certificate.
Non-microscopic		
1	**Clinical**	Diagnosis made before death, but without the benefit of any of the following (2–7)
2	**Clinical investigation**	To include all diagnostic techniques, including X-ray, endoscopy, imaging, ultrasound, exploratory surgery (e.g., laparotomy) and autopsy, without a tissue diagnosis.
4	**Specific tumour markers**	To include biochemical and/or immunological markers which are specific for a tumour site (Table 2).
Microscopic		
5	**Cytology**	Examination of cells whether from a primary or secondary site, including fluids aspirated using endoscopes or needles. Also to include the microscopic examination of peripheral blood films and trephine bone marrow aspirates.
6	**Histology of a metastasis**	Histological examination of tissue from a metastasis, including autopsy specimens.
7	**Histology of a primary tumour**	Histological examination of tissue from the primary tumour, however obtained, including all cutting techniques and bone marrow biopsies. Also to include autopsy specimens of a primary tumour.
9	**Unknown**	

Reference: IARC Recommendations on registry practices

- Clinical staging fields (eg. "clinical_tumour_staging_system", "clinical_stage_group" etc) should be a single instance at the time of primary diagnosis. If subsequent staging is

done on specimens taken later, these should be recorded in the pathological staging fields in the Specimen table.

- "clinical_stage_group" will be cross-validated against "clinical_tumour_staging_system". Refer to Tumour Staging Classifications documentation to ensure the correct stage group is submitted.
- If a version of AJCC is used in "clinical_tumour_staging_system", the "clinical_t_category", "clinical_n_category" and "clinical_m_category" fields will be cross-validated to ensure the correct TNM categories are used. For example, the term "Mx" cannot be submitted in "clinical_m_category" if AJCC 7th or 8th editions are used, since "Mx" was only allowed in AJCC 6th edition and prior.

8. Treatment

- Each treatment regimen must be assigned a unique "submitter_treatment_id".
- Additional supplementary treatment tables are required to be completed for each treatment indicated in the "treatment_type" field. If these are missing, clinical validation should fail. For example, if "treatment_type" includes "Chemotherapy", the supplementary "Chemotherapy" table is required.
- The "treatment_type" field accepts multiple values, so if a treatment regimen consists of two treatments given at the same time, these should be submitted together in the "treatment_type" field separated by a delimiter as indicated in the data dictionary, and the additional supplementary treatment tables should be submitted.
- The "response_to_treatment" field will be cross-validated against the "response_to_treatment_criteria_method" field. Refer to the Response to Treatment Criteria documentation to ensure the correct response_to_treatment term is submitted.
- The "is_primary_treatment" field should be used to indicate whether it is the first treatment given for the disease. It is often part of a standard set of treatments, such as surgery followed by chemotherapy and radiation. When used by itself, primary treatment is the one accepted as the best treatment. Also called first-line therapy, induction therapy, and primary therapy (Reference: NCI: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/primary-treatment)

How to submit a treatment regimen of Chemotherapy and Radiation therapy given at the same time

NOTE: Some columns are excluded for abbreviation

Submit the treatment regimen as one row in the Treatment table by submitting "treatment_type" = "Chemotherapy|Radiation therapy" and assign a unique "submitter_treatment_id" to it. ("Tr1" in this example).

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
MyDonor1	Tr1	PD01	Chemotherapy|Radiation therapy	2020-06	2020-12-29

The supplementary Chemotherapy and Radiation table are required. When submitting these two tables, submit the same "submitter_treatment_id" that was submitted in the Treatment table ("Tr1"). This indicates the two treatments were given at the same time.

Chemotherapy

submitter_donor_id	submitter_treatment_id	drug_reference_database	drug_reference_identifier	drug_name	chemotherapy_drug_dose_units	prescribed_cumulative_drug_dose
MyDonor1	Tr1	RxNorm	225852	Leucovorin calcium	mg/m2	3200
MyDonor1	Tr1	RxNorm	4492	fluorouracil	mg/m2	20000
MyDonor1	Tr1	RxNorm	32592	oxaliplatin	mg/m2	680

Radiation

submitter_donor_id	submitter_treatment_id	radiation_therapy_modality	radiation_therapy_type	radiation_therapy_fractions	radiation_therapy_dosage	anatomical_site_irradiated
MyDonor1	Tr1	External	Electron	60	500	Chest

How to submit primary treatment information consisting of neoadjuvant chemotherapy and surgery.

Submit the neoadjuvant chemotherapy and surgery as two separate entries in the Treatment table and indicate "is_primary_treatment" as "Yes".

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	is_primary_treatment	treatment_start_date	treatment_end_date	treatment_setting	treatment_intent	days_per_cycle	number_of_cycles	response_to_treatment_criteria_method	response_to_treatment
Donor 1	Tr1a	Pd01	Chemotherapy	Yes	2019-04	2019-08	Neoadjuvant	Curative	14	6	RECIST	Partial response
Donor 1	Tr1b	Pd01	Surgery	Yes	2019-09	2019-09	Induction	Curative			RECIST	Partial response

Submit the supplementary Chemotherapy and Surgery tables as well.

9. Chemotherapy

- The Chemotherapy table is required if "treatment_type" contains "Chemotherapy".
- If a chemotherapy treatment consists of multiple drugs, submit each drug as a separate row in this table and indicate the reference drug database it is found in, along with the drug reference identifier. The "drug_reference_database" and "drug_reference_identifier" will be used to validate the "drug_name" to ensure correct spelling.

drug_reference_database	**Drug_reference_identifier to submit**
RxNorm	RxCUI
PubChem	PubChem Compound Identifier (PubChem CID)
NCI Thesaurus	NCIt Code

Examples:
The drug_name "BMS-986315" is found in NCI Thesaurus (C176867)
The drug_name "Celastrol" is found in PubChem (PubChem CID 122724)

submitter_donor_id	submitter_treatment_id	drug_reference_database	drug_reference_identifier	drug_name	chemotherapy_ drug_dose_units	prescribed_cumulative_drug_dose
MyDonor1	Tr24	NCI Thesaurus	C176867	BMS-986315	mg/m2	60
MyDonor1	Tr25	PubChem	122724	Celastrol	mg/m2	20000

Example:
How to submit combination drugs such as "diclofenac/fluorouracil" which can be found in RxNorm with unique RXCUI assigned to it (1600381).

- Submit the two drugs as *two separate rows*, each with their own RXCUI (RXCUI "3355" for diclofenac and RXCUI "4492" for fluorouracil) . Both rows should be associated with the same "submitter_treatment_id" which indicates they were given together ("treatment_10" in this example). This enables tracking of the individual dosages for the two drugs.

submitter_donor_id	submitter_treatment_id	drug_reference_database	drug_reference_identifier	drug_name	prescribed_cumulative_drug_dose	chemotherapy_ drug_dose_units
MCH12	treatment_10	RxNorm	3355	diclofenac	35	mg/m2
MCH12	treatment_10	RxNorm	4492	fluorouracil	60	mg/m2

10. Immunotherapy

The Immunotherapy table is required if "treatment_type" contains "Immunotherapy". The drug names should be submitted the same way as described for the Chemotherapy table.

To denote immunotherapy drugs not found in any drug databases, such as engineered cellular therapy, or tailored personal therapies, submit "immunotherapy_type" = "Other immunomodulatory substances" and indicate the drug name in the "drug_name" field. The

"drug_reference_database" and "drug_reference_identifier" fields will not be required.

Example

submitter_donor_id	submitter_treatment_id	immunotherapy_type	drug_name	actual_cumulative_drug_dose	immunotherapy drug_dose_units
MTH15	treatment_12	Other immunomodulatory substances	SC935894	35	mg/m2

11. Radiation

The Radiation table is required if "treatment_type" contains "Radiation therapy".

How to indicate radiation boost
Submit information about the primary radiation treatment first ("Tr1") and the subsequent radiation boost treatment ("Tr2").

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
MyDonor1	Tr1	PD01	Radiation therapy	2020-06	2020-12
MyDonor1	Tr2	PD01	Radiation therapy	2020-12	2021-01

Radiation

submitter_donor_id	submitter_treatment_id	radiation_therapy_modality	radiation_therapy_type	radiation_therapy_fractions	radiation_therapy_dosage	anatomical_site_irradiated
MyDonor1	Tr1	External	Electron	60	500	Chest

If a radiation boost was given, submit as a new row in the Radiation table and submit "radiation_boost" = "Yes" and indicate the "submitter_treatment_id" of the primary radiation treatment the radiation boost treatment is linked to. In this example "Tr2" is the radiation boost treatment and is linked to the "Tr1" radiation treatment.

Radiation

submitter_donor_id	submitter_treatment_id	radiation_therapy_modality	radiation_therapy_type	radiation_therapy_fractions	radiation_therapy_dosage	anatomical_site_irradiated	radiation_boost	reference_radiation_treatment_id
MyDonor1	Tr1	External	Electron	60	500	Chest		
MyDonor1	Tr2	External	Electron	50	500	Head	Yes	Tr1

12. Surgery

The "Surgery" table is required if "treatment_type" contains "Surgery".

Refer to ARGO's Submitting Data in Surgery File documentation for guidelines on how to submit Surgery table.

Follow Up

A follow up is defined as any point of contact with a patient after primary diagnosis. To submit multiple follow ups for a single donor, assign a unique submitter_follow_up_id to each follow up and enter as a new row in the table.

If a follow up is related to a specific treatment that was entered in the Treatment table, you can reference the treatment in the submitter_treatment_id field.

13. Biomarker

Since biomarkers can be collected at different timepoints (eg. at the time of primary diagnosis, after treatment), each biomarker entry should be associated with a single clinical event (submitter_primary_diagnosis_id OR submitter_treatment_id OR submitter_follow_up_id OR submitter_specimen_id). If the biomarker is not associated with a clinical event, then you must indicate the date when the biomarker test was done.

You do not need to include every field in the Biomarker table since not all biomarkers will apply to donors in your cohort. Only include the biomarker fields that you have clinical data for. For example, if 100 donors in your cohort were only tested for ER and HER2 status, then you would only include fields relevant to ER and HER2 status in your table.

Example
Donor "donor1" had their PSA levels collected twice - once at the time of primary diagnosis, and again during a follow up. To indicate this, the data submitter will submit two entries in the Biomarker table for "donor1", each associated with an identifier for the clinical event the

biomarker was collected at. The clinical event identifiers (PD01 and F04) need to exist in the Primary Diagnosis and Follow Up table, respectively:

Primary Diagnosis:

submitter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
donor1	**PD100**	2020-08	C61

Follow Up:

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup
donor1	**F04**	2020-12	Stable

Biomarker:

submitter_donor_id	submitter_primary_diagnosis_id	submitter_followup_id	psa_levels
donor1	**PD100**		120
donor1		**F04**	56

14. Comorbidity

The Comorbidity table is optional. A donor's comorbidities are any medical conditions (e.g diabetes, prior cancer malignancies) that have existed or may occur during the clinical course of the donor who has the index disease under study. To submit multiple comorbidities for a single donor, submit multiple rows in the comorbidity file for this donor.

Example:

Donor "MTH5" has two comorbidities, a prior malignancy (ovarian cancer at age 23) and diabetes (diagnosed age 59)

submitter_donor_id	prior_malignancy	laterality_of_prior_malignancy	age_at_comorbidity_diagnosis	comorbidity_code	comorbidity_treatment_status
MTH5	Yes	Left	23	C57.8	No
MTH5			59	E10	Yes