The Division of Cancer Prevention

Minimum Data Set

Instructions and Guidelines

Version 5
January 8, 2026

Version Control

Date	Author(s)	Version
11/28/2012	Troy Budd	v2
6/11/2013	Troy Budd	v3
08/6/2013	DCP	V3.1
10/31/2013	Troy Budd	V4
01/14/2014	Troy Budd	V4.1
01/08/2026	Jennifer Pak / Troy Budd	V5

Questions and Comments

If you have any questions or comments regarding the Minimum Data Set (MDS) or the process for submitting your MDS submission to the Division of Cancer Prevention (DCP) via the secure File Transfer Protocol (sFTP) site, please contact the DCP Protocol Information Office (PIO) by phone 240-276-7130 or email at nci_dcp_pio@mail.nih.gov

Contents

1. MINIMUM DATA SET (MDS) INSTRUCTIONS

1.1 Overview

The MDS is a collection of specified administrative, participant demographic and adverse event data that serves as a primary source of clinical trials information about the NCI Division of Cancer Prevention (DCP) supported clinical trials. The MDS should be submitted for all DCP supported clinical trials for which DCP requires the MDS for routine data reporting. The MDS is routinely collected and reviewed by DCP, and routinely reported to NCI's Clinical Trials Reporting Program which maintains a comprehensive database of information on all NCI-supported interventional clinical trials.

1.2 Responsibility for Submission

The Lead Organization for a study is responsible for the MDS submissions. The Lead Organization is defined as the institution that is directly funded by DCP for the purpose of conducting the clinical trial.

1.3 Data Requirements

The MDS consists of those administrative, participant demographic and adverse event data elements specified by DCP. The definition, valid values and other details for each required MDS data element are consistent with the National Cancer Institute's Cancer Data Standards Registry and Repository (caDSR), and are listed in Appendix 1 (Minimum Data Set Table) of this document. Some data elements may be left blank when the data element is not applicable to be collected for a protocol.

The MDS includes the following data elements as defined in Appendix 1 including clarifications of the use of some elements for screening studies.

- DCP Protocol Number
- Submission Date
- Report Cut-off Date
- Current Trial Status
- Current Trial Status Date
- Name of Person Submitting the Data
- Submitter Telephone Number
- Submitter Email Address
- Participant Identifier
- Participant Zip Code
- Participant Country Code
- Participant Birth Date
- Participant Sex
- Participant Race
- Participant Ethnicity
- Informed Consent Date

- Screen 1 Date
- Screen 2 Date
- Registration Date
- Randomization Date
- Eligibility Status
- Participant Enrollment Date
- Registering Consortium (CTEP ID for the lead organization)
- Registering Institution
- Participant Method of Payment
- Treatment Assignment Code (TAC)
- Date Agent Started (or screening intervention start date for screening studies)
- Agent End Date (or screening intervention end date for screening studies)
- Off Study Date
- Off Study Reason
- Reason Off Study Other, Specify
- Adverse Event (AE) Verbatim Term
- MedDRA System Organ Class (SOC)
- CTCAE Term
- Other, Specify (to be used for MedDRA SOC= Injury, poisoning and procedural complications and CTCAE Term= Injury, poisoning and procedural complications - Other, specify)
- AE Grade
- AE Attribution
- Reported as SAE?
- Event Onset Date
- Event End Date
- Dropped Due to AE?
- Outcome

1.4 MDS Submission Schedule

MDS submissions are due monthly. **Files should be successfully submitted by the 10th of each month.**

Each submission should include all required data available on the last day of the month preceding the monthly MDS due date (see the chart in Section 1.5). This date is defined as the Report Cut-off Date. **All MDS submissions must be cumulative, i.e. all data collected for a study from its approval date to the specified report cut-off date must be included in each submission.**

1.5 Initial MDS Submission Dates

The first MDS submission is due the 10th day of the second month after DCP Final Study Approval.

Guideline for initial submission:

Date of DCP Approval	First MDS Due Date	Reporting Period
January	March 10	Approval – February 28 (29)
February	April 10	Approval – March 31
March	May 10	Approval – April 30
April	June 10	Approval – May 31
May	July 10	Approval – June 30
June	August 10	Approval – July 31
July	September 10	Approval – August 31
August	October 10	Approval – September 30
September	November 10	Approval – October 31
October	December 10	Approval – November 30
November	January 10	Approval – December 31
December	February 10	Approval – January 31

1.6 MDS Submission Period

MDS submissions are required monthly until a study reaches a status of 'Complete' or 'Administratively Complete'. These statuses are defined as follows:
Complete -Study has been closed to accrual, all participants have completed treatment or intervention and the study has met its primary objectives.
Administratively Complete -The protocol has been completed prematurely (e.g., due to poor accrual, insufficient drug supply, IND closure).

A MDS submission is required if a study has been closed to accrual and intervention but the primary objective has not been met, or if a study has been approved but has not yet been activated.

1.7 Methods of MDS Submission

All MDS submissions must be submitted electronically in a standardized format using the specified DCP secure File Transfer Protocol (sFTP) site. ***Paper submissions will not be accepted.***
Each electronic file must contain only the required cumulative data for a single study. Electronic files that contain data for multiple studies will not be accepted.

Please refer to Section 2 for the specific electronic file format requirements.

DCP MDS sFTP site

A DCP MDS sFTP site is established for the secure transfer of MDS submissions between the submitting organization and DCP. To ensure the security and integrity of the data included in all MDS submissions, an account with a username and password will be created by DCP for each person at the organization designated by the Lead Organization to submit MDS submissions.

1.8 Account Creation

For the creation of an account to access the DCP MDS sFTP site, please contact IMS Support Team by email at **nci-mds@imsweb.com** . They will need the following information to create an account.

- User's full name

- User's organization
- Email address
- Telephone number

Note: DCP will grant a maximum of two accounts per Lead Organization.

2. FILE FORMAT INSTRUCTIONS

2.1 Introduction

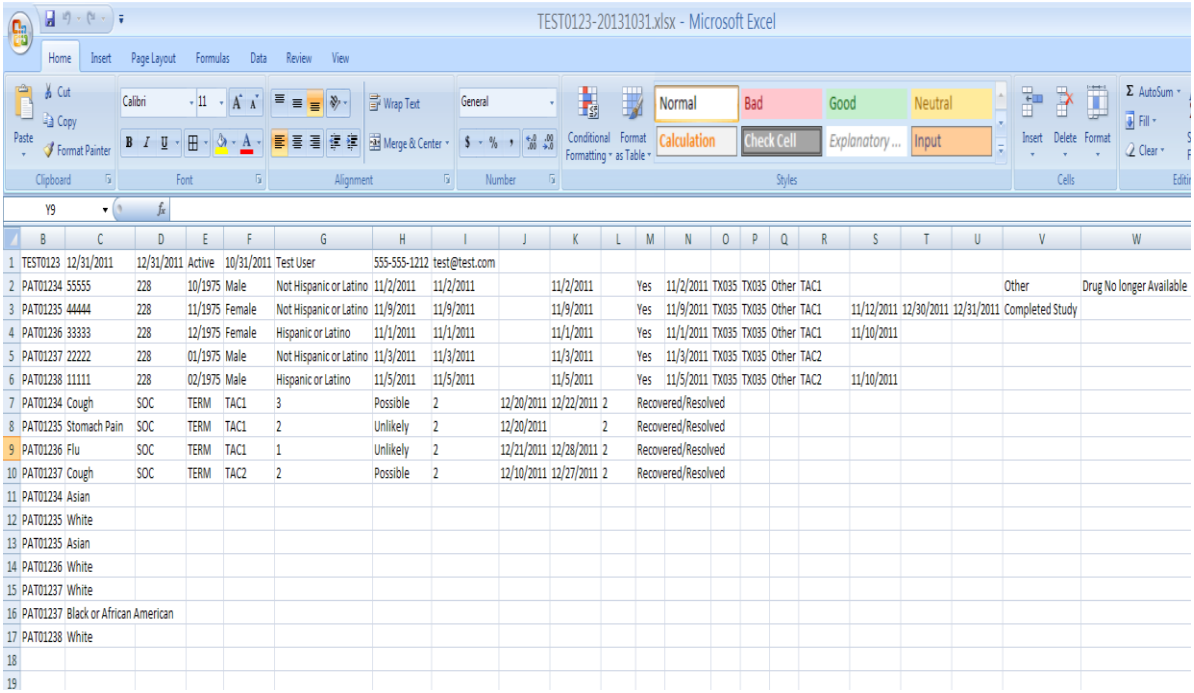
The MDS submission is designed to populate the DCP database from a single comma-delimited file that is electronically submitted to DCP.

A sample of the required file format is shown below.

CVS file opened in Notepad (comma delimited file)

CVS file opened in Excel

Do not double-click a CSV file to view it in Excel because Excel will make assumptions about your data and may format it incorrectly. Instead, you should open Excel, go to the Data tab, and choose “From Text/CSV” to import the file.



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	TEST0123	12/31/2011	12/31/2011	Active	10/31/2011	Test User	555-555-1212	test@test.com														
2	PAT01234	55555	228	10/1975	Male	Not Hispanic or Latino	11/2/2011	11/2/2011		11/2/2011		Yes	11/2/2011	TX035	TX035	Other	TAC1				Other	Drug No longer Available
3	PAT01235	44444	228	11/1975	Female	Not Hispanic or Latino	11/9/2011	11/9/2011		11/9/2011		Yes	11/9/2011	TX035	TX035	Other	TAC1	11/12/2011	12/30/2011	12/31/2011	Completed Study	
4	PAT01236	33333	228	12/1975	Female	Hispanic or Latino	11/1/2011	11/1/2011		11/1/2011		Yes	11/1/2011	TX035	TX035	Other	TAC1	11/10/2011				
5	PAT01237	22222	228	01/1975	Male	Not Hispanic or Latino	11/3/2011	11/3/2011		11/3/2011		Yes	11/3/2011	TX035	TX035	Other	TAC2					
6	PAT01238	11111	228	02/1975	Male	Hispanic or Latino	11/5/2011	11/5/2011		11/5/2011		Yes	11/5/2011	TX035	TX035	Other	TAC2	11/10/2011				
7	PAT01234	Cough	SOC	TERM	TAC1	3	Possible	2	12/20/2011	12/22/2011	2	Recovered/Resolved										
8	PAT01235	Stomach Pain	SOC	TERM	TAC1	2	Unlikely	2	12/20/2011		2	Recovered/Resolved										
9	PAT01236	Flu	SOC	TERM	TAC1	1	Unlikely	2	12/21/2011	12/28/2011	2	Recovered/Resolved										
10	PAT01237	Cough	SOC	TERM	TAC2	2	Possible	2	12/10/2011	12/27/2011	2	Recovered/Resolved										
11	PAT01234	Asian																				
12	PAT01235	White																				
13	PAT01235	Asian																				
14	PAT01236	White																				
15	PAT01237	White																				
16	PAT01237	Black or African American																				
17	PAT01238	White																				
18																						
19																						

Appendix I - Minimum Data Set Table

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
Protocol	DCP Protocol Number	The unique alphanumeric identifier assigned to a protocol by the Division of Cancer Prevention (DCP)		Character	35,1	977 v3.0
Protocol	Submission Date	The date on which the report is to be submitted	MM/DD/YYYY	Date	10	
Protocol	Report Cut-off Date	The end date of the reporting interval	MM/DD/YYYY	Date	10	2992 v4.0
Protocol	Current Trial Status	The current status of a clinical study	Active Administratively Complete Approved Closed to Accrual Closed to Accrual and Intervention Complete Temporarily Closed to Accrual Temporarily Closed to Accrual and Intervention Withdrawn	Character	50	2518475 v1.0
Protocol	Current Trial Status Date	The date that the current trial status became effective	MM/DD/YYYY	Date	10	2200228 v1.0
Protocol	Name of Person Submitting the Data	The legal name of the person who is submitting the data		Character	87	2006163 v1.0
Protocol	Submitter Telephone Number	The telephone number where the person completing/submitting the report can be reached		Character	20,7	2200276 v1.0
Protocol	Submitter Email Address	The email where the person completing/submitting the report can be reached		Character	100	2200278 v1.0

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
Participant; Race; AE	Participant Identifier	The unique numeric or alphanumeric identification assigned to a participant in a clinical trial or research study		Character	20	2003301 v4.0
Participant	Participant Zip Code	The string of characters used to identify the five-digit zone improvement plan (ZIP) code and the four-digit extension code (if available) that represents the geographic segment that is a subunit of the ZIP code, assigned by the U.S. Postal Service to a geographic location to facilitate mail delivery; **or the postal zone specific to the country, other than the U.S., where the mail is delivered**		Character	15	2179606 v2.0
Participant	Participant Country Code	The code that represents the country where the addressee is located		Alphanumeric	3,3	2179605 v1.0
Participant	Participant Birth Date	The month and year on which the person was born	MM/YYYY	Date	7	793 v5.1
Participant	Participant Sex	Text designations that identify sex. Sex is described as the assemblage of properties that distinguish people on the basis of their societal roles [Identification of sex is based upon self-report and may come from a form, questionnaire, interview, etc.]	Female Male Unknown Unspecified	Character	13,4	2200604 v3.0
Participant	Participant Ethnicity	The text for reporting information about ethnicity based on the Office of Management and Budget (OMB) categories	Hispanic or Latino Not Hispanic or Latino Unknown Not Reported	Character	22,7	2192217 v2.0
Participant	Informed Consent Date	The date on which the patient/participant/legal representative agrees OR disagrees to participation in a protocol, treatment, or other activity by signing an informed consent document	MM/DD/YYYY	Date	10	656 v4.0

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
Participant	Screen 1 Date	Date participant completes Screen 1. This is the step to determine study eligibility for the potential participant. For studies where the intervention is a screening procedure, this is not the date the participant undergoes the first screening procedure.	MM/DD/YYYY	Date	10	2184691 v1.0
Participant	Screen 2 Date	Date participant completes Screen 2	MM/DD/YYYY	Date	10	2184691 v1.0
Participant	Registration Date	The date the patient was enrolled on the protocol	MM/DD/YYYY	Date	10	2171 v4.0
Participant	Randomization Date	Date of a process used in therapeutic trials or other research endeavors for allocating experimental subjects, human or animal, between treatment and control groups, or among treatment groups	MM/DD/YYYY	Date	8,4	2182072 v1.0
Participant	Eligibility Status	The yes/no indicator that asks the investigator to stipulate whether the participant is eligible for inclusion on this protocol	Yes No	Character	7,2	1235 v4.0
Participant	Participant Enrollment Date	The date the participant is accepted into the study. The study site may also be notified to the treatment arm and Study Participant Identifier on this date	MM/DD/YYYY	Date	10	2746541 v1.0
Participant	Registering Consortium	The designation of a consortium that will be officially recorded as the registering consortium for the study		Character	5	2813153 v1.0
Participant	Registering Institution	Code that uniquely identifies the institution where the research participant was registered in a clinical trial		Character	10	2003307 v4.0

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
Participant	Participant Method of Payment	Text term for an entity, organization, government, corporation, health plan sponsor, or any other financial agent who pays a healthcare provider for the healthcare service rendered to a person or reimburses the cost of the healthcare service	Private Insurance Medicaid Medicaid and Medicare Military Sponsored (including CHAMPUS & TriCare) Veterans Sponsored No Means of Payment (No Insurance) Medicare Medicare and Private Insurance Self-Pay (No Insurance) Managed Care/Medicare State Supplemental Health Insurance Military or Veterans Sponsored, NOS Other Unknown	Character	50	2865130 v1.0
Participant, AE	Treatment Assignment Code (TAC)	A coded value representing a treatment assigned to be uniformly administered to a group of study subjects for separate statistical analysis		Character	10,1	1967 v4.0
Participant	Date Agent Started	The start date for the administration of the agent intervention. For screening studies, this is the date the participant undergoes the first screening procedure.	MM/DD/YYYY	Date	8	3028744 v1.0
Participant	Agent End Date	The end date for the administration of the agent intervention. For screening studies, this is the date the participant undergoes the last screening procedure. If there is only one screening procedure, the “Agent End Date” will be the same date as the “Date Agent Started”.	MM/DD/YYYY	Date	8	3028746 v1.0
Participant	Off Study Date	The date when the participant is removed from the protocol, i.e., is not being followed and will not be retreated	MM/DD/YYYY	Date	8,4	2003605 v3.0

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
Participant	Off Study Reason	Choice of reasons for removing a participant from a clinical trial	Adverse Event Death Disease Progression Lost to follow-up Other, specify Participant Withdrawal Participant Refused Follow-up Physician Decision Protocol Defined Follow-up Completed Protocol Violation Study Complete Ineligible	Character	50	2979313 v1.0
Participant	Reason Off Study Other, Specify	The text that describes the reason the participant went off study		Character	200	2182613 v1.0
Race	Participant Race	The text for reporting information about race based on the Office of Management and Budget (OMB) categories	American Indian or Alaska Native Asian Black or African American Native Hawaiian or Other Pacific Islander Not Reported Unknown White	Character	41,5	2192199 v1.0
AE	Adverse Event (AE) Verbatim Term	The text that describes the adverse event word for word as described by the participant		Alphanumeric	200	2188132 v1.0
AE	MedDRA System Organ Class (SOC)	Text term to represent the highest level of a terminology distinguished by anatomical or physiological system, etiology, or purpose, and referencing an international medical terminology (Medical Dictionary for Regulatory Activities) version 12.0, designed to support the classification, retrieval, presentation, and communication of medical information throughout the medical product regulatory cycle.		Character	80	2943864 v1.0
AE	CTCAE Term	Text that represents the Common Terminology Criteria for Adverse Events lowest level term name for an adverse event.		Character	84	3125302 v1.1

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
AE	Other, Specify	Text to specify with the Adverse Events of Special Interest (AESI). For vaccine/immunologic trials whose protocol specifies to use this "Other, Specify" AE data element, fill in using the following AESI: "Post-immunization reaction (systemic)" OR "Post-immunization reaction (local)".		Character	100	2433262 v1.0
AE	AE Grade	Numeric representation of the intensity/severity of an unfavorable and unintended sign (including an abnormal laboratory finding), symptom, syndrome, or disease, temporally associated with the use of a medical product or procedure, regardless of whether or not it is considered related to the product or procedure (attribution of unrelated, unlikely, possible, probable or definite)	0 = Absent Adverse Event 1 = Mild Adverse Event 2 = Moderate Adverse Event 3 = Severe Adverse Event 4 = Life-threatening Adverse Event 5 = Death Related to Adverse Event	Character	1,1	2944515 v1.0
AE	AE Attribution	Relation of the causality between the treatment modality and the specific adverse event	Unrelated Unlikely Possible Probable Definite	Character	10,1	1285 v3.0
AE	Reported as SAE?	The code representing whether the event was reported as a Serious Adverse Event	1 = Yes 2 = No	Character	1,1	2182930 v1.0
AE	Event Onset Date	The date on which the adverse event was first evident	MM/DD/YYYY	Date	8,4	2744993 v1.0
AE	Event End Date	The last or final date of an adverse event, described using a date or a text response such as Ongoing or Unknown	MM/DD/YYYY	Date	8	2189843 v1.0
AE	Dropped due to AE?	Did the participant stop participation due to AE	1 = Yes 2 = No	Character	1	2683534 v1.0

Collection Table	Data Element	Definition	Permissible	Data Type	Field Size – max and min, if appropriate	caDSR Public ID
AE	Outcome	The final status of the participant related to the adverse event	Recovered/Resolved Recovering/Resolving Not Recovered/Not Resolved Recovered/Resolved with Sequelae Fatal Unknown	Character	33	2746517 v1.0

Appendix II - GLOSSARY

AE – Adverse Event

CaDSR - Cancer Data Standards Registry and Repository

CTCAE - Common Terminology Criteria for Adverse Events

DCP – Division of Cancer Prevention

FTP – File Transfer Protocol

IND - Investigational New Drug Application

MDS – Minimum Data Set

NCI – National Cancer Institute

OMB – Office of Management and Budget

PIO – Protocol Information Office

SAE – Serious Adverse Event

sFTP – Secure File Transfer Protocol

SOC - System Organ Class

TAC – Treatment Assignment Code