UNOS
UNITED NETWORK FOR ORGAN SHARING

Matching organs. Saving lives.

700 North 4th Street, Richmond, VA 23219
tel: 804-782-4800
fax: 804-782-4816
www.unos.org

Brian M. Shepard
Executive Director & CEO

TO: UNOS STAR File Researcher

FROM: UNOS Research Department

RE: Standard STAR Files and Documentation

DATE: October 2021

In response to your request, enclosed is a copy of the UNOS Standard Transplant Analysis and Research File. This file is commonly referenced as the "National UNOS STAR file".

Most answers about the OPTN data collection fields can be found by reading this document.

Some OPTN data collection fields will not be provided due to the data release policies. If a field is not found within the data files or documentation, it may not be available publicly.

A data dictionary within the file STAR File Documentation.xls is included in accompanying folders. This is the guide to using the STAR files and includes extensive documentation about each dataset and each variable. **See Appendix A at the end of this letter for more information.**

The most current OPTN data collection forms are provided online at https://unos.org/technology/. These forms will change with the addition, removal, and/or modification of fields, lookup values, and labels.

For details about OPTN data collection, please use the following links:
https://www.unos.org/data/
https://www.unos.org/data/technology-for-transplantation/

We recommend that you start by reviewing all documentation, to familiarize yourself with the variables and formats that are most pertinent to your analysis. Keep in mind that many variables have only been collected for the period of time specified within the documentation. As such, it is important to include only those variables collected for the entire period the analysis needs.

Since March 2020, for accuracy and completeness of the data, we have removed the 2-month lag used in reporting that previous STAR Files included. The data files created based on data as of October 1, 2021 will include data for registrations added and/or transplants occurring through September 30, 2021.

Each main organ-specific file (i.e. KIDPAN_DATA, LIVER_DATA, etc.) contains information on all waiting list registrations and transplants of that organ type that have been listed or performed in the U.S. and reported to the OPTN since October 1, 1987. It includes both deceased and living donor transplants. There is one record per waiting list registration/transplant event, and each record includes the most recent follow-up information (including patient and graft survival) reported to the OPTN as of the date the file was created. If a patient was listed for a transplant, but was removed prior to transplant or is still waiting, all the transplant information for that patient

is null (i.e. there are waiting list records in the dataset with no transplant information). Similarly, if a patient received a living donor transplant, and was never on the waiting list, all of the waiting list specific information for that patient is null (i.e. there are transplant records in the dataset with no waiting list information). Waiting list registrations can be selected by choosing records where WL_ID_CODE is not null, and transplants performed can be selected by choosing records where TRR_ID_CODE is not null. If there was a waiting list registration that resulted in a transplant event, neither WL_ID_CODE nor TRR_ID_CODE will be null. Additionally, we've introduced new variables; VAL_DT_TCR and VAL_DT_TRR, providing information about when the Transplant Candidate Registration and Transplant Recipient Registration was validated. These variables might be null for more recent registrations/transplants.

The follow-up portion of the STAR File (i.e. KIDPAN_FOLLOWUP_DATA, LIVER_FOLLOWUP_DATA, etc.) contains one record per follow-up per transplant event. Therefore, in most cases you will find multiple records per transplant. For instance, if a patient received a transplant in January 2001, the graft has not failed, and the patient has not been reported lost to follow up, you should find multiple follow-up records (a record at 6 months, and one at each year after 2001) with the same transplant identification number (TRR_ID_CODE). In the same example, if the patient received a simultaneous kidney-pancreas transplant, you should find follow-up records recorded for kidney alone and follow-up records recorded for pancreas alone until January 2003, and then a single kidney-pancreas follow-up record for each year onward. On follow-ups generated since January 2003, a single kidney-pancreas follow-up is generated in place of the kidney alone and pancreas alone forms. These more recent data are found in the kidney-pancreas follow-up file. Similar to the reporting for registration and transplant information, the data files created in October 2021 will include follow-up data where the follow-up forms are expected to be completed prior to October 1, 2021. The variable for linking all the follow-up data to the main datasets is TRR_ID_CODE. Additionally, we've introduced a new variable; VAL_DT_TRF, providing information about when the Transplant Recipient Follow-up was validated. This variable might be null for more recent follow-ups.

Similar to the follow-up information explained above, additional waiting list information can be linked by WL_ID_CODE to the main dataset, additional information at the time of transplant can be linked by TRR_ID_CODE to the main dataset, and additional follow-up information can be linked by TRR_ID_CODE to the main dataset and by TRR_FOL_ID_CODE to the follow-up dataset.

These files do not include any patient or transplant hospital identifiers. However, there is an encrypted patient identification number (PT_CODE), unique to each patient that allows you to track the patient through multiple waiting list and transplant events. Also there are encrypted transplant and donor hospital/OPO identifiers for tracking purpose
(CTR_CODE, OPO_CTR_CODE, LISTING_CTR_CODE, INIT_OPO_CTR_CODE,
END_OPO_CTR_CODE, RECOV_FACILITY_CODE).

The organ-specific transplant files include most of the pertinent donor information collected on the corresponding organ donor (living or deceased) such as donor age, donor gender, donor history of hypertension, etc. Additionally, the dataset “DECEASED_DONOR_DATA” contains more detailed information on all deceased donors in the U.S. (one record per donor) where at least one organ was transplanted or recovered for transplant and reported to the OPTN since October 1, 1987. This allows the researcher to obtain additional information about each transplanted organ, and also to examine other issues, such as deceased donor organ recovery

and utilization. The data can be linked to the corresponding transplants using the field DONOR_ID. Additionally, we've introduced a new variable; VAL_DT_DDR, providing information about when the Deceased Donor Registration was validated. This variable might be null for more recent registrations.

Similarly, the file “LIVING_DONOR_DATA” contains information on all living donors recovered in the U.S. since October 1, 1987. For living donors recovered prior to October 25, 1999, there are limited medical and demographic fields available. The data can be linked to the corresponding transplants using the field DONOR_ID. Additionally, we've introduced a new variable; VAL_DT_LDR, providing information about when the Living Donor Registration was validated. This variable might be null for more recent registrations.

The file “LIVING_DONOR_FOLLOWUP_DATA” contains one record per follow-up visit for living donors recovered since October 25, 1999. Prior to that date, living donor follow-up information was not collected. For living donors recovered since then, there will be the potential for a six month and one year follow-up on donors recovered from October 25, 1999 to February 29, 2008 with the additional two year follow-up form added for donors recovered since March 1, 2008. Additionally, we've introduced a new variable; VAL_DT_LDF, providing information about when the Living Donor Follow-up was validated. This variable might be null for more recent follow-ups.

The end-user assumes all responsibility for analyses performed with these data. We encourage you to consider the time period during which each field was collected when performing any analysis. Deaths and graft failures may be reported on interim follow-ups, but patients who are alive are more likely to be reported at follow-ups submitted on the anniversary of the transplant. Therefore, performing an analysis too early following the end of the follow-up period may lead to artificially low estimates of survival as deaths may have already been reported but living patients may not have reached their anniversary yet. For instance, if your analysis time point is survival at 1 year for transplants performed between January 2011 and December 2012, we would recommend that the analysis be performed only after the beginning of March 2014 (December 2012 + 1 year to reach analysis end point + 2 month lag).

The data are provided in three file formats: Tab delimited text files, SAS datasets, and SAS Export to STATA format. If you have access to SAS software, the SAS dataset files located within the folder named “SAS Datasets” are recommended. The SAS catalogs, sas7bcat, contain the SAS formats. If you are using SPSS, it is recommended to import the SAS dataset files into SPSS. The lookup values and SAS formats are also located within subfolders as tab delimited text files under the main folder “CODE DICTIONARY – FORMATS”.

When first using the data, you will note that all of the folders (one for each of the three formats described in the paragraph above) have been compressed (zipped) to fit onto a single folder. If you are using WINZIP or PKZIP, you can access the data directly in the appropriate folder structure. If you are using SECUREZIP, you will need to “extract” the data prior to using it in order to maintain the desired folder structure. The SECUREZIP wizard can walk you through that process. It is recommended to extract all data before use.

Included in the data files provided are the results of linkage between the OPTN database and additional external sources of deaths. If the patient was reported as deceased from OPTN data or by a verified external source, the COMPOSITE_DEATH_DATE will be populated with that date of death. This additional death information is included in the calculated time in days of patient

survival, PTIME. The variables, PX_STAT and PX_STAT_DATE, provide the most recent patient information for that transplant event as reported by the transplant center that followed the patient at the time of STAR file copy date. The COMPOSITE_DEATH_DATE may be different from PX_STAT_DATE. For example, if the last patient status from the center was reported as January 1, 2014 and the patient was reported as alive, but the COMPOSITE_DEATH_DATE reports the patient death date on June 30, 2014, then all additional variables on the files will reflect the patient status reported as alive on January 1, 2014 according to the OPTN data and *not* the external sources death date provided. However, the variables COMPOSITE_DEATH_DATE, PSTATUS, and PTIME will all reflect the death date reported for patient as deceased as verified by external sources.

If you have any questions, please email STARFile@unos.org (Email is preferred method of contact as the mailbox is monitored daily by Research staff.)

Special Considerations When Using OPTN Data in a Manuscript (including Abstracts) or Presentation

1. **Citing Data Obtained Through the Data Request System.** Please cite in the text or graphic presentation, the date and source of the data. The data extract date is mentioned on first page of this document. For example, "Based on OPTN data as of October 12, 2012." Additionally, please add the following statement as an acknowledgment: "This work was supported in part by Health Resources and Services Administration contract 234-2005-370011C. The content is the responsibility of the authors alone and does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government."
2. **IRB Approval.** If these data are to be used as part of a presentation or manuscript (including abstracts), you must check with your institution to see whether IRB approval (expedited or otherwise) is required.

Please note: The UNOS Research Department requests that any UNOS staff member who makes a significant intellectual contribution to a study abstract, presentation, or manuscript be offered the opportunity to be included as an author. UNOS staff may not be listed as study authors without obtaining written permission from the appropriate staff member(s).

APPENDIX A

The STAR File Documentation.xls located within the **IMPORTANT DOCUMENTATION** folder is the data file documentation or data dictionary for using the files, variables, and lookup values.

The following is provided within the STAR File Documentation.xls file:

- a detailed description of the dataset and variable names and labels
- where the data have been collected (FORM), unless calculated
- the variable collection time period (VAR START DATE, VAR END DATE)
- the SAS format assigned to the coded variable (SAS Analysis Format)
- any comments or additional information for the particular variable (COMMENT)

The first worksheet, "**Contents**", provides links and a description of the contents for each file.

The second worksheet, "**File Formats**", provides a description of the file formats within each folder on disc. See "Recommended use" for the suggested data format.

Every disc contains the following folders:

- Delimited Text File (.DAT extension)
  - Tab delimited file
- SAS Dataset (.sas7bdat extension)
  - SAS raw file (v9.03.01M2)
- SAS Export to STATA (.DTA extension)
  - SAS file exported to STATA

The third worksheet, "**Main Documentation**", provides information about the main folders Delimited Text File, SAS Dataset, and SAS Export to STATA) The column "FILE NAME" lists every file available and hyperlinks to the corresponding worksheet within the workbook. Each corresponding worksheet includes the variables, data collection details, and SAS formats or lookup values specific to the individual file(s).

For example, the first file listed, "DECEASED_DONOR_DATA", links to the worksheet "DECEASED_DONOR_DATA". On the linked worksheet "DECEASED_DONOR_DATA", the variable name, description, form section, variable start date, variable end date, data type, SAS analysis format (or lookup values), SAS system data length, SAS format, SAS variable label, and any additional comments with information about the field are provided.

The fourth worksheet, "**Formats (Lookup Values)**", provides information about the folder named "CODE DICTIONARY – FORMATS". All format lookup values are located on both the worksheet "**Formats (Lookup Values**)" and in the folder named "CODE DICTIONARY – FORMATS". The "FILE NAME" column links to the worksheets with the workbook including the same information found within the folder (data field formatted value, SAS format, data type, and data field value).

For example, the KIDPAN_DATA formats are included in the KIDPAN_FORMATS_FLATFILE worksheet and file. If the variable on the KIDPAN_DATA file lists a SAS analysis format, the lookup value is located within the KIDPAN_FORMATS_FLATFILE. The variable ABO_MAT lists the SAS analysis format $ABOMAT. The data field values of ' ', '1', '2','3' match to formatted values of 'Unknown', 'Identical', 'Compatible', and 'Incompatible'.

The remaining worksheets include all the individual files and formats linked in other worksheets.