Atherosclerosis Risk in Communities Study

Cohort Surveillance Recurrent Events Derived Variable Dictionary

Cohort Surveillance Recurrent Events Derived Variable Dictionary

The recurrent events dataset contains every major CVD event that occurred in the ARIC cohort: MI, CHD death, stroke, and heart failure (both adjudication-based and discharge code-based), as well as all cardiac procedures and deaths. The dataset contains cohort IDs for only the participants that had events of one of these types, so you will need to merge the dataset back with, for instance, visit 1 data in order to perform a survival analysis of baseline factors and time to CVD event.

The philosophy when designing the dataset was to be "user-centric," not "user-friendly"[1] by giving users a transparent dataset that gives them the freedom and responsibility to do whatever analysis they wish. As such, there are some things to know about the dataset that would be helpful to know before you dive in:

1. For a given event time, every type of event that could have occurred at that event time is listed as a separate row. For example, if a participant died of CHD death on January 1, 2004, there will be a row for the event "CHD death" and a row for the event "DEATH". This construction allows users to concentrate on any outcome that they wish, without fear that they are leaving out events that could have been classified multiple ways.

2. Related to #1, we included HF events according to discharge codes and MMCC adjudication. One consequence of this choice is that the count of the number of events where type_of_event = "HF" does not represent the number of hospitalizations for HF among ARIC participants. If an HF hospitalization was identified by both discharge codes and MMCC review, then there will be one row for the discharge code event and one row for the MMCC review event, with the dates of the events being identical. This format allows users to use whichever definition they wish, or even a combined outcome.

3. The appropriate censoring time, CENSDAT7, is already applied to all outcomes except for death (not CHD death). Since we receive deaths from NDI as well as ARIC surveillance, there should not be a bias in follow up time by disregarding CENSDAT7 when analyzing total mortality.

Many resources are available online to learn about analyzing correlated event times, whether the correlation is due to the type of event (e.g., "competing risks") or multiple events within the same individual.

Table of Contents

1.   Classification Variables

1.1 TYPE_OF_EVENT

Purpose

To identify type of CVD event or death among cohort participants

Description

TYPE_OF_EVENT classifies the type of event that happened for a given event time. The types are

- CHD death (definite fatal CHD or definite fatal MI),
- MI (definite or probable MI),
- stroke (definite or probable subarachnoid hemorrhage, brain hemorrhage, thrombotic infarction, or non-carotid embolic infarction),
- heart failure (428.X or I50.X for discharge codes, definite or probable acute decompensated heart failure for MMCC review),
- cardiac procedures (36.x)
- and death.

1.2 TYPE_OF_MI

Purpose

To identify the type of MI when TYPE_OF_EVENT = "MI"

Description

TYPE_OF_MI takes on the values "STEMI" or "NSTEMI" if the event is an MI and is missing otherwise.

1.3 STROKE_SUBTYPE

Purpose

To identify the type of stroke when TYPE_OF_EVENT = "STROKE"

Description

STROKE_SUBTYPE specifies whether the stroke was a definite or probable subarachnoid hemorrhage, brain hemorrhage, thrombotic infarction, or non-carotid embolic infarction if the event is a stroke and is missing otherwise.

1.4 HF_SUBTYPE

Purpose

To identify the type of HF when TYPE_OF_EVENT = "HF" and SOURCE_OF_HF_EVENT = "ADJUDICATION"

Description

HF_SUBTYPE specifies whether the HF event was a definite or was a probable acute decompensated HF.

1.5 SOURCE_OF_HF_EVENT

Purpose

To identify the source of identification for an HF event

Description

SOURCE_OF_HF_EVENT has the value "discharge codes" when the event was identified by a 428 or I50 on the CEL or DTH form, while the value "adjudication" refers to an event identified and classified using cohort surveillance, which began in 2005.

1.6 SOURCE_OF_DEATH_INFORMATION

Purpose

To identify the source of information for a death

Description

SOURCE_OF_DEATH_INFORMATION takes the value "ARIC" when the event was identified through cohort follow-up and the value "NDI" when the death was identified by searching the National Death Index.

2. Time Variables

2.1 TIME_SINCE_VISIT 1

Purpose

To create a variable of time to event for survival analysis

Description

TIME_SINCE_VISIT_1 reports the time in days between the participant's Visit 1 in the ARIC study and the current event of interest.

2.2 CENSDAT7_FollowUpDays

Purpose

To determine the follow-up time in days since visit 1 to censoring date for follow up time of cohort participants for events identified through cohort surveillance, cohort follow-up or linkage with registries.

Description

CENSDAT7 is the date of the last known status for all cohort participants for follow-up of any type of event, except death. It is determined by the last date of contact with the participant or proxy: date of visit 2, date of visit 3, date of visit 4, date of visit 5, date of visit 6 or from the interview date of the annual or semi-annual follow up. Annual and semi-annual follow-up interviews are counted only if hospitalization information was collected. If the participant has died and the date of last contact occurred within a year prior to death, then follow-up is censored at the death date. Otherwise, follow-up is censored at the last contact, as previously described.

This variable should be used to censor time-to-event analysis of events identified through cohort surveillance (CEL), cohort follow-up (SAFU, AFU) or linkage with registries. Follow-up for time-to-death analysis may be continued beyond last contact; see the description for DATED17.

CENSDAT7 is a modification of CENSDAT6, which additionally allowed for follow-up through last event date for events identified through cohort surveillance, or death date as identified through cohort surveillance, NDI linkage or annual follow-up. However, there are two problems with this definition:

- Censoring on last event date includes participants lost to follow-up in the risk set through the date of last hospitalization if the hospitalization occurred in an ARIC community hospital. Participants lost to follow-up with no hospitalizations or hospitalizations outside of ARIC catchment will be excluded from the risk set. Once a participant is lost to follow-up they would continue in the risk set only if a hospitalization is identified, i.e. they contribute to the denominator only if they contribute to the numerator. This differential follow-up may bias time-to-event analysis.

- Censoring on death date implies that hospitalization information is available until time of death. This will not be the case for participants with deaths identified through the NDI search who have moved out of the community and are not participating in cohort follow-up.

For these reasons, all time-to-event analysis (except time to death) should be censored at last contact as defined by CENSDAT7 and in corresponding variables with the prefix C7_. Events identified after CENSDAT7, including those identified through linkage with registries, should be excluded from analysis.

Algorithm

(a). If max(V2DATE21, V3DATE31, V4DATE41, V5DATE51, V6DATE61, LAST_COMPLETE_INTERVIEW)>= 31DEC17 then C7_CENSDAT7=31DEC17.
(b). If not above then C7_CENSDAT7=max(V2DATE21, V3DATE31, V4DATE41, V5DATE51, V6DATE61, LAST_COMPLETE INTERVIEW).

Related Variables

V2DATE21 (cohort visit 2 date), V3DATE31 (cohort visit 3 date), V4DATE41 (cohort visit 4 date), V5DATE51 (cohort visit 5 date), V6DATE61 (cohort visit 6 date), V7DATE71 (cohort visit 7 date), LAST_COMPLETE_INTERVIEW(Follow up Last Contact Date).

2.3 CENSDAT7_Year

Purpose

To determine the year of the censoring date for follow up time of cohort participants for events identified through cohort surveillance, cohort follow-up or linkage with registries.

3. Identification Variables

3.1 COHORT_ID

Purpose
>To identify a cohort member

Description
>The original cohort ID who experienced the current event of interest

3.2 ID

Purpose
>To uniquely identify an event for a given participant

Description
>This variable serves as a unique identifier of an event identified through ARIC's cohort study. There may be multiple rows for a given surveillance ID if the event can be classified as more than one event type.