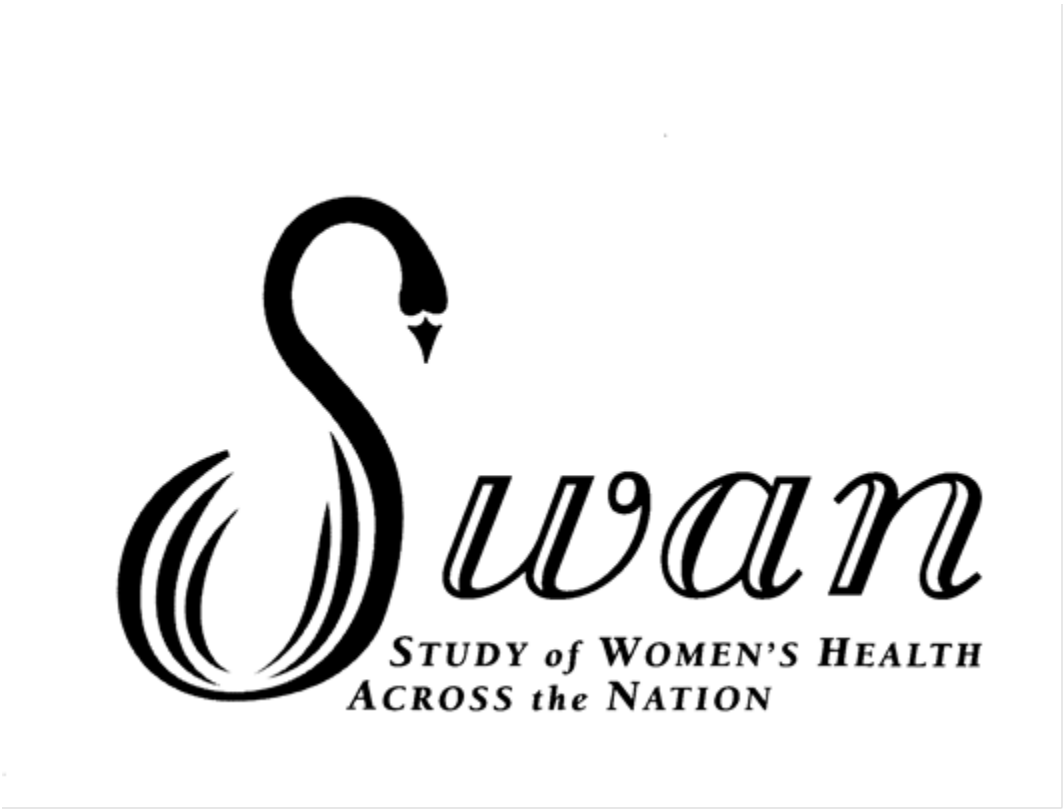


Longitudinal Mortality Dataset

Baseline through Visit 17

CODEBOOK

ARCHIVED DATASET 2026

Dataset file niamort2026

PAGE INTENTIONALLY BLANK

DOCUMENTATION FOR THE SWAN LONGITUDINAL MORTALITY DATASET

This dataset contains information on mortality in the SWAN cohort up-to-date through the end of Visit 17. It combines the previously created death certificate and vital status datasets with the newly obtained NDI data.

SWAN study sites obtained death certificates from family members of participants throughout SWAN follow-up. Death certificates (DC) were de-identified by the study sites and sent to the Coordinating Center, where data were extracted based on the Death Certificate Data Extraction Form. Relevant information collected includes date of death, place of death, underlying cause of death, presence of specific conditions on any line of the death certificate, and manner of death. If a death certificate was not obtained for a participant, they are still included in the data set but have missing information for all DC fields.

In SWAN-Aging (Visit 17), mortality status was assessed using the National Death Index for the first time. The NDI is a centralized database of death record information from 1979 through 2022 managed, through the National Center for Health Statistics (NCHS) in state vital statistics offices. The SWAN data core in coordination with each study site submitted NDI search applications for participants who have been lost to follow-up. For each death identified, the date of death and cause of death codes were collected.

Who is included in the public use dataset:

The dataset contains information for the 3,302 women from the seven clinical sites participating in the SWAN longitudinal study, **320** of whom have died. Of these 320 participants, death certificates have been obtained for 264, and NDI reports have been obtained for 82. The sites include 11, 12, 13, 14, 15, 16 and 17.

How this codebook is constructed:

This documentation section contains in depth explanation of the longitudinal mortality dataset. A list describing the variables used to create this dataset is also provided, along with the questionnaires on which many of the variables are based. Note that none of those variables are available in this dataset (they can be retrieved from each visit's dataset).

The assigned participant ID has been replaced with a randomly generated ARCHID in order to protect participant privacy. The baseline interview date is denoted as day 0 and is used as the basis for all other dates. Variables describing the race/ethnicity of participants (RACE) was added from the Screener dataset.

Missing data coding:

Original missing codes (-1: not applicable, -7: refused, -8: don't know, -9: missing) have been recoded to SAS missing codes (.B: not applicable, .D: refused, .C: don't know, and .A: missing).

NOTE: Please be aware that there was for site 16 women a gap of 2-3 visits between Visits 06 and 09. This might affect the accuracy of the created variables.

Decoding NDI Reports:

The data obtained from the NDI search included: date of death, underlying cause of death, cause recodes, entity-axis conditions, and record-axis conditions. While date of death is straight forward, the remaining four pieces of data are alphanumeric codes that represent specific ICD-10-CM (International Classification of Diseases, Tenth Revision, Clinical Modification) diagnoses. As such, extensive decoding was required to convert these into analysis ready information. The NDI search initially contained 595 SWAN participants who were lost to follow-up, from which 82 matching reports were found.

The main source used for the decoding process was the Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses, v2023.1, created by the Healthcare Cost and Utilization Project (HCUP). The purpose of this system is to match the ICD-10-CM alphanumeric codes to their clinical diagnoses, as well as match them to a clinically meaningful category. The ICD-10-CM contains upwards of 70,000 individual codes, thus the CCSR breaks them down into 530 aggregate categories to simplify analyses. These categories have been further generalized in this dataset in order to maintain the privacy of SWAN participants.

Regarding Death Certificates:

According to death certification guidelines, the last line of the death certificate should have the underlying cause of death listed. However, not all certifying physicians follow these guidelines, resulting in the most proximal cause of death (for example, respiratory failure), listed at the bottom. This then leads to a loss of information if the death certificate was not coded according to standards.

Some issues identified while cleaning the death certificate data are as follows:

- Death certificates were not obtained for **56** IDs confirmed dead by other sources (36 via STAT form and 20 via NDI).
    - For **36** of these IDs, we sought out their death certificates during the Visit 17 period, and were not able to obtain them. These participants were included in the dataset with all death certificate data as .B type missing, except for GETCERT and DCOMPDAY.
    - **One** of these IDs had participated in Visit 17, and we learned of their death later during the period of the visit. Since this participant died after V17, their death certificate will not be requested until the next visit. All death certificate data for this participant is set to .A type missing.
    - For **19** of these IDs, the participants had already been deactivated at prior visits. We therefore only learned of their deaths through the NDI search, and had not included them in the list of IDs to request death certificates for. Attempts will be made during the next visit to obtain their death certificates, but for now, all death certificate data is set to .A type missing.
- There were **20** IDs where date of death (DOD) from death certificates did not match the original SWAN study site-determined date of death. These could be due to error by the physician certifying death, site error, the mode by which date of death was determined by SWAN investigators (for example, if the death was originally reported by a family member as just a month and year, the 15th of that month is used as a placeholder for the actual date), or differences in determining circumstances of death. Each of these 20 cases use the date of death (converted to days from baseline interview) reported on the death certificate for the day of mortality status variable.
- There were **16** IDs where we obtained date of death only from the death certificate, and did not have a reported death date on a STAT form. In all of these cases, the participant had previously been deactivated, and we received notice of their death via death certificate during Visit 17.
- There were **17** IDs where we obtained DOD only from a STAT form, and could not obtain a death certificate or NDI report to confirm.
- There were **2** IDs where we were not able to obtain a death certificate and did not have a reported DOD on a STAT form, nor were they matched in the NDI search conducted during Visit 17. In both of these cases, a family member reported their death on the V17 STAT form, but did not provide DOD.

There was **1** ID where we originally obtained DOD from a STAT form and could not obtain a death certificate, but we were able to match them in the NDI search and the date found there differed from our records. In this case, DOD was revised to match the date in the NDI report.

Sources:

1. World Health Organization; Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, based on the recommendations of the Ninth Revision Conference, 1975; Geneva; 1977.
2. National Center for Health Statistics. *National Death Index User's Guide.* Hyattsville, MD. 2024.
3. National Center for Health Statistics. *Instruction Manual: ICD-10 Cause-of-Death Lists for Tabulating Mortality Statistics (Updated March 2011 to include WHO updates to ICD-10 for data year 2011).* Hyattsville, MD. 2011.
4. Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. Healthcare Cost and Utilization Project (HCUP). November 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp
5. International Classification of Diseases Tenth Revision (ICD-10), Sixth Edition, 2019. Geneva: World Health Organization; 2019. License: CC BY-ND 3.0 IGO. https://icd.who.int/browse10/2019/en.

Variables included in the dataset:

Variable	Label	Codes
ARCHID	Participant ID	
RACE	Race/Ethnicity	1 = Black 2 = Chinese/Chinese American 3 = Japanese/Japanese American 4 = White Non-Hispanic 5 = Hispanic
MORTSTAT	Mortality Status	0 = Alive 1 = Deceased
MORTSRC	Mortality Information Source	VISIT = Annual Visit Participation (Interview or other form) STAT = Annual follow-up status form MYHR = Mid-Year Health Report MENS = Menstrual Calendar DC = Death Certificate NDI = NDI Report
MORTDAY	Day of Mortality Status	
DCOMPDAY	Day of Death Certificate Form Completion	
GETCERT	Death certificate obtained?	1 = No 2 = Yes
DPLACE	Place of death	1 = Hospital 2 = Hospice 3 = Nursing home/LTC 4 = Decedent's Home 5 = Other Residence 6 = Ambulance 7 = Other
DCAUSE	COD available from Cause of Death section of DC?	1 = No 2 = Yes
CODCV	Heart disease listed on any line of DC	1 = No 2 = Yes
CODHF	Heart failure listed on any line of DC	1 = No 2 = Yes
CODMI	Heart attack/MI listed on any line of DC	1 = No 2 = Yes
CODCVA	Stroke listed on any line of DC	1 = No 2 = Yes
CODPNEUM	Pneumonia listed on any line of DC	1 = No 2 = Yes
CODCANC	Cancer listed on any line of DC	1 = No 2 = Yes
CODDIAB	Diabetes listed on any line of DC	1 = No 2 = Yes
CODCOPD	COPD listed on any line of DC	1 = No 2 = Yes
CODKIDNE	Kidney failure listed on any line of DC	1 = No 2 = Yes
CODLIVER	Liver failure or cirrhosis listed on any line of DC	1 = No 2 = Yes
CODROAD	Road injury listed on any line of DC	1 = No 2 = Yes
CODFALL	Fall listed on any line of DC	1 = No 2 = Yes
DINJURY	Did DC indicate that COD was an injury?	1 = No 2 = Yes
NDIREPORT	NDI Report Available?	1 = No 2 = Yes
NCOMPDAY	Day of NDI Report Form Completion	

Variable	Label	Codes
NDICAT	Categorical Underlying Cause of Death from NDI Report	[Text]

DATA COLLECTION FORMS

 # = Variable excluded from public use datafile
~ A randomly generated ID will be provided that is different from the original ID
† This date is given in days since the initial baseline interview, which is day zero.

Date Data Entered / Initials _____ Date Verified / Initials _____

DEATH CERTIFICATE DATA EXTRACTION FORM

Study of Women's Health Across the Nation

SECTION A. GENERAL INFORMATION

AFFIX ID LABEL HERE

A1. PARTICIPANT ID: **ARCHID~**

A2. FORM VERSION: **#FORMV_DC** 02/16/2021

A3. DATE FORM COMPLETED: **DCOMPDAY†** ___ ___ / ___ ___ / ___ ___ ___ ___
M M D D Y Y Y Y

A4. FORM COMPLETED BY (INITIALS): ___ ___ ___ **#INITS**

A5. PARTICIPANT'S DOB: **#DOB** ___ ___ / ___ ___ / 1 9 ___ ___
M M D D Y Y Y Y

A6. Was a death certificate obtained? **GETCERT**

NO ..... 1 **END FORM**
YES 2

SECTION B. DEATH CERTIFICATE INFORMATION FORM

B1. DATE OF DEATH (FROM DEATH CERTIFICATE): **#DODMON/#DODDAY/#DODYR**

MORTDAY†

____ ____	/	____ ____	/	____ ____ ____ ____
M M		D D		Y Y Y Y

B2. PLACE OF DEATH (if possible, identify type of facility, rather than the name or address) **DPLACE**

Hospital 1
Hospice facility 2
Nursing Home/Long-term care facility 3
Decedent's Home 4
Other Residence 5
Ambulance 6
Other 7
Specify **#DPLACES**_____

B3. Were any causes of death available from the "Cause of Death" section of the death certificate? **DCAUSE**

NO ..... 1 **(B4)**
YES 2

B3a. What is the cause of death listed on the **last non-blank line** of the "Cause of Death" section of the death certificate? (NOT the "contributing causes" listed below the "Cause of Death" section) **#COD**

_____

B3b. Were any of the following causes listed on **any line** of the "Cause of Death" section of the death certificate? (NOT including the "contributing causes" section below the "Cause of Death" section):

		NO	YES	
a.	Heart disease	1	2	CODCV
b.	Heart failure	1	2	CODHF
c.	Heart attack/Myocardial Infarction	1	2	CODMI
d.	Stroke	1	2	CODCVA
e.	Pneumonia	1	2	CODPNEUM
f.	Cancer	1	2	CODCANC
g.	Diabetes	1	2	CODDIAB
h.	Chronic Obstructive Pulmonary Disease (COPD)	1	2	CODCOPD
i.	Kidney failure	1	2	CODKIDNE
j.	Liver failure/cirrhosis	1	2	CODLIVER
k.	Road Injury	1	2	CODROAD
l.	Fall	1	2	CODFALL

B4. Did the death certificate indicate that the cause of death was an injury? **DINJURY**

NO 1
YES 2

B5. Was a “Manner of Death” available from the death certificate? **#DMANNER**

NO ..... 1 **END FORM**
YES 2

B5a. What was the “Manner of Death” listed on the death certificate: **#MODDC**

Natural 1
Accident 2
Suicide 3
Homicide 4
Could not be determined 5
Pending investigation 6
None of the above 7

Date Data Entered / Initials _____ *Date Verified / Initials* _____

NDI DATA ENTRY (NDAF)

Study of Women's Health Across the Nation

SECTION A. GENERAL INFORMATION

AFFIX ID LABEL HERE

A1. PARTICIPANT ID: **ARCHID~**

A2. SWAN STUDY VISIT # 17 **#VISIT**

A3. FORM VERSION: 04/17/2023 **#FORMV_NDI**

A4. DATE ENTRY COMPLETED: ____ ____ / ____ ____ / ____ ____ ____ ____ **NCOMPDAY†**
M M D D Y Y Y Y

A5. ENTRY STAFF INITIALS: _____ _____ _____ **#INITS**

A6. DID SITE OBTAIN A MATCH FOR THE PARTICIPANT? **#MATCHOBT**

NO……1 (END FORM)
YES……2 (B1)

SECTION B. DEATH INFORMATION FROM REPORT FILE

B1. DATE OF DEATH ____ ____ / ____ ____ / ____ ____ ____ ____ **MORTDAY†**
M M D D Y Y Y Y

SECTION C. DEATH INFORMATION FROM PRTCAUSE FILE

C1. DID SITE OBTAIN A CAUSE/PRTCAUSE FILE FOR THE PARTICIPANT? **#PRTCAUSE**

NO……..1 **(END FORM)**
YES……2 **(C2)**

C2. UNDERLYING CAUSE OF DEATH CODE: ____ ____ ____ ____ **#DEATHC**

C3. 282 OR 358 CAUSE RECODE: ____ ____ ____ ____ ____ **#DCR358**

C4. 72 OR 113 CAUSE RECODE: ____ ____ ____ **#DCR113**

ENTITY ACCESS CONDITIONS

C5. ENTITY ACCESS CONDITION 1 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA1**

C6. ENTITY ACCESS CONDITION 2 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA2**

C7. ENTITY ACCESS CONDITION 3 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA3**

C8. ENTITY ACCESS CONDITION 4 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA4**

C9. ENTITY ACCESS CONDITION 5 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA5**

C10. ENTITY ACCESS CONDITION 6 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA6**

C11. ENTITY ACCESS CONDITION 7 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA7**

C12. ENTITY ACCESS CONDITION 8 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA8**

C13. ENTITY ACCESS CONDITION 9 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA9**

C14. ENTITY ACCESS CONDITION 10 ____ ____ ____ ____ ____ ____ ____ **#DEATHEA10**

RECORD ACCESS CONDITIONS

C15. RECORD ACCESS CONDITION 1 ____ ____ ____ ____ ____ **#DEATHRA1**

C16. RECORD ACCESS CONDITION 2 ____ ____ ____ ____ ____ **#DEATHRA2**

C17. RECORD ACCESS CONDITION 3 ____ ____ ____ ____ ____ **#DEATHRA3**

C18. RECORD ACCESS CONDITION 4 ____ ____ ____ ____ ____ **#DEATHRA4**

C19. RECORD ACCESS CONDITION 5 ____ ____ ____ ____ ____ **#DEATHRA5**

C20. RECORD ACCESS CONDITION 6 ____ ____ ____ ____ ____ **#DEATHRA6**

C21. RECORD ACCESS CONDITION 7 ____ ____ ____ ____ ____ **#DEATHRA7**

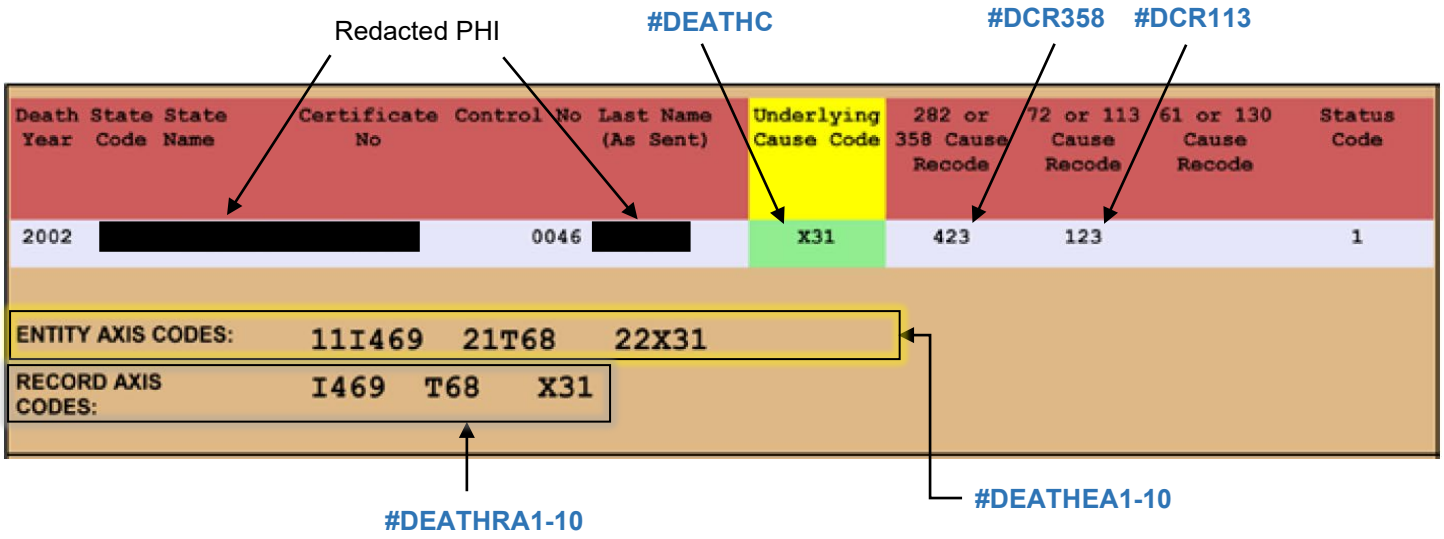
C22. RECORD ACCESS CONDITION 8 ____ ____ ____ ____ ____ **#DEATHRA8**

C23. RECORD ACCESS CONDITION 9 ____ ____ ____ ____ ____ **#DEATHRA9**

C24. RECORD ACCESS CONDITION 10 ____ ____ ____ ____ ____ **#DEATHRA10**

Example NDI Report

Note: This example comes from the National Death Index User's Guide, and does not pertain to any SWAN participants.



- DEATHC is recoded into NDICAT