Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research:

Pilot Linkage of National Death Index+ to Commercially and Publicly insured Populations

Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration

Enhancing Data Resources for Studying Patterns and Correlates of Mortality in Patient-Centered Outcomes Research: Pilot Linkage of NDI+ to Commercially and Publicly Insured Populations

Executive Summary

Mortality is a common outcome of interest in pharmacoepidemiologic studies assessing medication safety and effectiveness. Distributed networks of electronic health care databases commonly used to conduct patient-centered outcomes research (e.g., electronic health records, health insurance claims data) typically capture medically attended deaths during enrollment but often do not systematically capture out-of-hospital deaths or cause of death information. The capability to link distributed data networks to the National Death Index (NDI) will augment many types of patient-centered outcomes research, including medication safety surveillance, predictive risk modeling, and comparative effectiveness research.

This report summarizes standardized and reusable methods developed for attaining death and cause of death information from the Centers for Disease Control and Prevention's NDI in multi-site studies. We piloted this process in a study designed to assess the risk of all-cause mortality and specific causes of death following select antiarrhythmic medications. The developed linkage process avoids sharing identifiable patient information among study sites or with a central coordinating center. This approach promotes data sharing by protecting patient privacy, data security, as well as any other confidential information.

In this study report, we provide a guide with step-by-step recommendations for multi-site studies that require linkage with NDI data, along with lessons learned (**Attachment 1**). Chapter 1 describes a tested administrative workflow to facilitate efficient, coordinated, multi-center Institutional Review Board (IRB) review and approval for linking health plan data with NDI data, which includes death and cause of death information. This chapter also provides recommendations for completing a successful NDI application along with lessons learned that may help future studies navigate the process more efficiently. Chapters 2-5 describe a standardized and reusable distributed technical process for efficiently attaining and analyzing death and cause of death information from the NDI across multiple health plan databases without sharing protected health information between health plans or with the Coordinating Center. The appendix includes a detailed description of lessons learned that may be applicable to future studies. To implement the technical process described in Chapters 2-5, we created multiple programming packages. The technical specifications for the prototype programs that were developed and implemented within the pilot study could be adapted and reused in future studies are included in **Attachment 2**.

Attachment 3 describes the pilot study analysis we used to develop the distributed NDI linkage process. As a test case, we identified a cohort of antiarrhythmic medication users and non-users in six health plans in years 2000-2017. The users and non-users were 1:1 matched on age, sex, and calendar time.

Patients with any of the following to the NDI for linkage: (a) health plan recorded death during the exposure episode plus 365 days, (b) potential death, or (c) incomplete follow-up (i.e., health plan data ended before patients could be followed for the entire exposure episode plus 365 days). Potential death was defined as health plan disenrollment between cohort entry and cohort exist plus 365 days, without reenrollment or medication utilization more than 60 days post-disenrollment. At five health

plans that ascertain death with state death records and other sources we examined concordance of death identified through NDI linkage compared to health plan recorded death data as reference.

In total, distributed programs identified 60,785 antiarrhythmic users and 60,785 matched non-users at six health plans, and we submitted 33,403 patients to the NDI for linkage (26% were patients with health plan recorded death, and 74% were patients identified with potential death or incomplete follow-up). The NDI returned at least one potential match for 67% of patients submitted to the NDI. A program for selecting the best NDI match classified 40% of patients submitted to the NDI as dead (99% of patients with health plan recorded death and 20% of patients with potential death or incomplete follow-up were classified as dead with the NDI linkage).

At a subset of health plans that ascertain death with state death records and other sources, high concordance was observed between health plan recorded death and NDI-identified death (96% agreement; Cohen's Kappa=0.919). The developed approach standardized the NDI linkage across health plans, improved capture of mortality outside of health plan data, reduced misclassification of death within the study, and therefore supports the utility of NDI linkage within mortality studies.

In summary, this study developed a standardized and reusable distributed NDI linkage process which can be leveraged by other studies. Throughout the study report, we describe the methods, recommendations, and lessons learned from the pilot study and summarize flexible options that might be helpful to future studies. Overall, the developed NDI linkage process may be particularly useful for multi-center studies using a common data model, as it prioritizes efficiency, consistency, and practicality. We highlight practical considerations and lessons learned such as utilizing a centralized IRB review process, early communication with NDI advisors to understand the requirements and expectations, and applying standardized technical approaches to help study sites navigate through the process. Future work will further test and refine the developed linkage approach, with the goal of improving the ability to answer mortality-related research questions within multi-center distributed data networks.

Attachments:

1. Study report: Distributed Process for Attaining Death and Cause of Death Information from the National Death Index (NDI) in Multi-Center Studies

2. Supplemental Appendix: Distributed Program Specifications Implemented in a Pilot Study

3. Pilot study results presented at the 36th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE) in 2020: A standardized and reusable method to link multiple distributed health plan databases to the National Death Index

Disclaimer: This report reflects the views of the authors and should not be construed to represent FDA's views or policies.

Attachment 1
Distributed Processes for Attaining Death and Cause of Death Information from the National Death Index (NDI) in Multi-Center Studies

Lead Study Site and Coordinating Center at Harvard Pilgrim Health Care Institute:
Candace Fuller, Andrew Petrone, Robert Rosofsky, Laura Shockro, Christine Draper, Michael Dizinno, Katelyn King, Sengwee Toh

Vanderbilt University Medical Center:
Marie Griffin, Margaret Adgent, Jae Young Min, Judy King, Jean Bruce, James Morrow

HealthPartners Institute:
Pamala Pawloski, Erin Schwartz, Teri Defor, Sunita Thapa

Kaiser Permanente Colorado:
Ingrid Binswanger, Daniel Jaynes, Karen Glenn

Kaiser Permanente Northwest:
Jennifer Kuntz, Chris Catlin, Phillip Crawford

Kaiser Permanente Washington:
Denise Boudreau, Monica Fujii, Vina Graham, Ron Johnson

Aetna, A CVS Health Company:
Cheryl McMahill-Walraven, Smita Bhatia, Annemarie Kline, Carla Brennan, Djeneba Audrey Djibo,

University of Pennsylvania:
Charles Leonard, Samantha Soprano

University of Iowa:
Ryan Carnahan

Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration:
Wei Hua, Andrew Mosholder, Sarah Dutcher, Carla Walls, Denise Jones, Rita Noel, Robert Ball

December 23, 2020

Document Distributed Processes for Attaining Death and Cause of Death Information from the National Death Index (NDI) in Multi-Center Studies

Table of Contents

Introduction: Working in Distributed Data Networks

Many multi-center research studies are conducted within distributed data networks of electronic health care databases, in which individual study sites or health plans maintain physical and operational control over their data behind their respective firewalls. For example, the US Food and Drug Administration (FDA)'s Sentinel System utilizes a distributed data network approach to monitor post-market safety of approved medical products.[1–3]

This guide summarizes a distributed process we developed for attaining death and cause of death information from the National Death Index (NDI) in multi-center studies.[4] The NDI is a self-supporting service within the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention[5] that includes a centralized database of death record information compiled from the vital statistics offices of states and other jurisdictions. The NDI provides death information including death date and death certificate number (referred to as NDI data), as well as cause of death from death certificates (referred to as NDI Plus or NDI+ data) upon request. A distributed approach to conducting linkages to the NDI is especially appealing given the patient-level protected health information (e.g., name, date of birth, social security number) required to attain death and cause of death information from the NDI. A distributed NDI linkage process and procedure in multi-center research settings would promote efficiency and standardization while minimizing concerns regarding patient privacy, data security, or proprietary interests.[1–3]

This project piloted a distributed process for attaining death and cause of death information in multi-center studies through a use case designed to examine the associations between select antiarrhythmic medications and all-cause mortality and specific causes of death.[4] The project was led and coordinated by the Harvard Pilgrim Health Care Institute (HPHCI), which worked closely with the FDA and participating health plans. Six health plans – Healthagen, LLC. (formerly Aetna, Inc.), HealthPartners, Kaiser Permanente Colorado, Kaiser Permanente Northwest, Kaiser Permanente Washington, and Vanderbilt University Medical Center (which provides access to Tennessee Medicaid data) – participated in this project. They represent a diverse group of health plans, including national insurers, regional health plans, and integrated delivery systems, and cover both commercial and public insurance programs. Although the project leveraged the Sentinel infrastructure and was built on the successful collaboration among participating health plans in Sentinel, it was conducted outside of Sentinel.

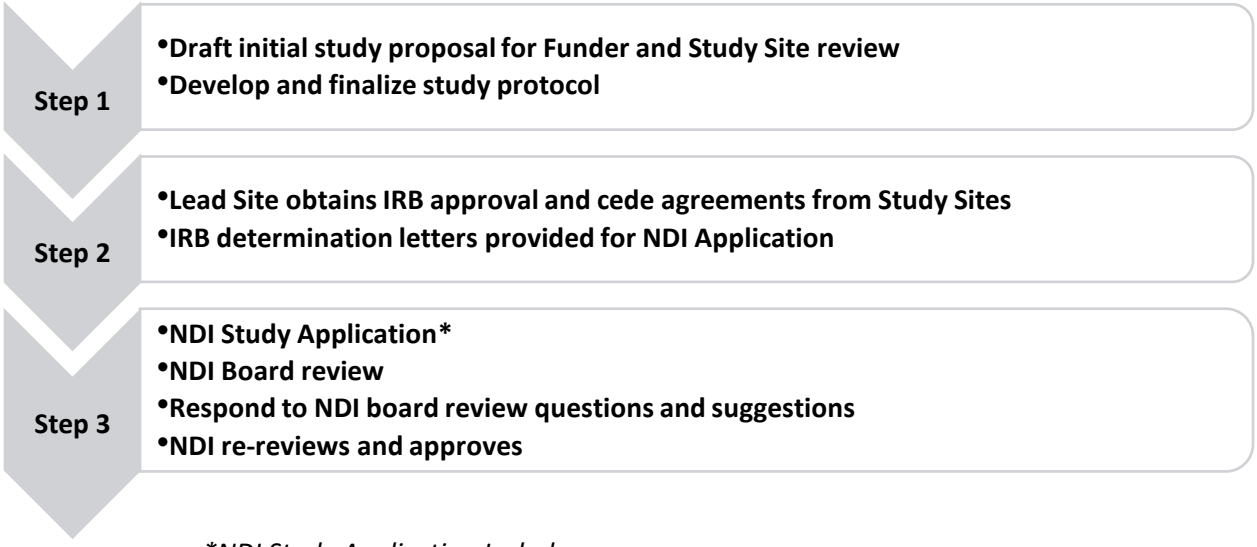
In this guide, we describe a tested administrative workflow to facilitate efficient, coordinated, multi-center IRB review and approval for linking health plan data with NDI+ data in accordance with the revised Common Rule.[6] We provide recommendations for completing a successful NDI application, along with lessons learned that may help future studies navigate the process more efficiently. We also describe a standardized and reusable distributed technical process for efficiently attaining and analyzing death and cause of death information from the NDI across multiple health plan databases without sharing protected health information between health plans or with the Coordinating Center. We provide technical specifications for the prototype programs we developed and implemented within our pilot study that could be adapted and reused in future studies. Additionally, we provide considerations for determining which patients to submit to the NDI for matching. Throughout this guide, we describe lessons learned through developing and testing these NDI linkage methods and summarize any flexible options that might be helpful to future studies. We developed these NDI linkage processes with the goal of improving the ability to answer mortality-related research questions within multi-center distributed data network studies.

Chapter 1: Administrative Workflow

This chapter provides an overview of the piloted administrative workflow required to support simultaneous linkage of multiple health plan databases with death data from the NDI. We developed this chapter to provide future multi-center research studies with a guide to administrative processes required to conduct an NDI linkage project. We include guidance on working with the NDI to initiate and complete an NDI application, and efficiently soliciting Institutional Review Board (IRB) approval in multi-center research settings. We also provide process recommendations based on lessons learned while piloting this administrative workflow. Technical aspects required for linking multiple study site databases to NDI data are described in **Chapters 2-5**.

This administrative workflow was piloted in a multi-center research study funded by the FDA. The required administrative tasks are outlined in **Figure 1**. **Table 1** describes the roles and responsibilities utilized during the pilot of the administrative workflow to facilitate use by other multi-center studies.

Figure 1. Administrative steps required to conduct linkage to NDI data in a multi-center research setting



**NDI Study Application Includes:*

1. *Protocol Specifics*
2. *IRB Determination Letters*
3. *Executed Confidentiality Agreements*
4. *Data Management & Disposition Plan*

Table 1: Piloted Study Team Administrative Roles and Responsibilities

Role Name	Members	Role Description	Responsibilities
Project Lead Teams	Sponsor: *FDA* Coordinating Center: *HPHCI* Principal Investigator	Develops study protocol, oversees scientific aspects of study in collaboration with Scientific Experts and Study Sites	• Oversees administrative workflow needed to support NDI linkage activities • Develops study protocol for IRB review that describes study population and case submission for NDI linkage • Selects lead IRB site for Central IRB review • Works with the Coordinating Center and Study Sites to determine availability of data elements required for the study, (i.e., data inventory) and identify Study Sites needs
Coordinating Center	Coordinating Center: *HPHCI*	Manages administrative processes for achieving linkage of multiple study databases to NDI data	• Manages IRB processes, opens initial IRB application with the central IRB, collaborates with Study Sites and Sponsor to ensure all steps required for Central IRB review and approval are achieved • Develops the NDI Application, shares and reviews Study Site applications for consistency, and execution of any required legal agreements by all parties to link to the NDI • Manages communications between the NDI and the Study Sites to ensure NDI data request is approved
Study Sites	Health Plans	Collaborates with the Project Lead Team and Coordinating Center on protocol and administrative tasks	• Manages internal administrative processes and scientific review at each Study Site, as well as agreement for site IRB to cede to a Central IRB • Works with the Coordinating Center to complete the NDI application, and execute any required legal documents • Works with Project Lead Teams/Coordinating Center to identify required data elements required for the study (i.e., data inventory) identify any internal administrative processes

Abbreviations: FDA: US Food and Drug Administration; HPHCI: Harvard Pilgrim Health Care Institute; NDI: National Death Index; IRB: Institutional Review Board

I. Initiating a study requesting death and cause of death information from the NDI

A draft protocol or proposal outlining the study question is required to complete an NDI application and initiate the administrative workflow needed to achieve approvals to link multiple study databases to NDI data. The NDI currently requires any study intending to use NDI data to undergo IRB review, and will not approve a data application until IRB approvals are received from all sites participating in the research. As study activities cannot proceed prior to IRB approval, the Project Lead Team (**Table 1**) should provide a draft protocol for IRB review in advance of study initiation. If a draft protocol is not available, the study

should include a time for protocol development in the study timeline. We recommend allowing at least four months for protocol development, based on the multi-study site pilot experience. The Project Lead Team may wish to conduct a feasibility analysis to ensure adequate sample sizes are available to support the study question. A feasibility analysis often can be conducted as a 'Preparatory to Research 'request [45 CFR 164.512(i)(1)(ii)].[7] The Project Lead Team should consider whether a feasibility analysis is needed and discuss with Study Sites regarding the possibility of preparatory to research feasibility analysis. If a feasibility analysis is required, it is important to account for any additional time required to conduct the analysis and review results in the study timeline. A draft protocol or proposal for initial IRB review at all Study Sites at a minimum must include:

- Study aims, including description of study population, and Study Sites
- Description of exposures, covariates, and outcomes of interest (e.g., all-cause mortality, death from specific causes) and general study analysis plan
- Planned use and transfer of the data collected for the study, data security measures, and data disposition plans (including description of end of study data disposition plans)

While a draft protocol or proposal is sufficient for initiating required administrative workflow, any changes and final study protocols must be submitted for IRB review before changes are implemented.

II. Soliciting Central IRB approval

Although use of a Central IRB is not required by the NDI, we chose to proceed with a Central IRB during the pilot of administrative workflows because the Coordinating Center and Study Sites determined it was the most efficient path forward. We found two major steps were necessary to efficiently achieve Central IRB approval for multi-center research studies requiring NDI data linkage. An initial IRB application must be submitted to open the study, and then the study sites can proceed with steps to cede to a Central IRB.

A. Initial IRB review

The Project Lead Team should submit an initial application to the appropriate overseeing IRB or lead IRB of record (i.e., if a Commercial IRB is utilized) prior to initiating a centralized IRB review process. The Coordinating Center handled this task in the pilot project, but it is important to recognize selecting a lead IRB for a multi-center study is a collaborative effort. We found utilizing the IRB affiliated with the Coordinating Center as the lead IRB site for the centralized IRB process created efficiencies but required approval from the Coordinating Center's affiliated IRB and participating Study Site IRBs. Institutions should consider institutional and ethical review responsibilities and be comfortable with relying on with the chosen lead IRB for the centralized IRB process.

To open an initial IRB application in the pilot project, the Coordinating Center was required to submit:

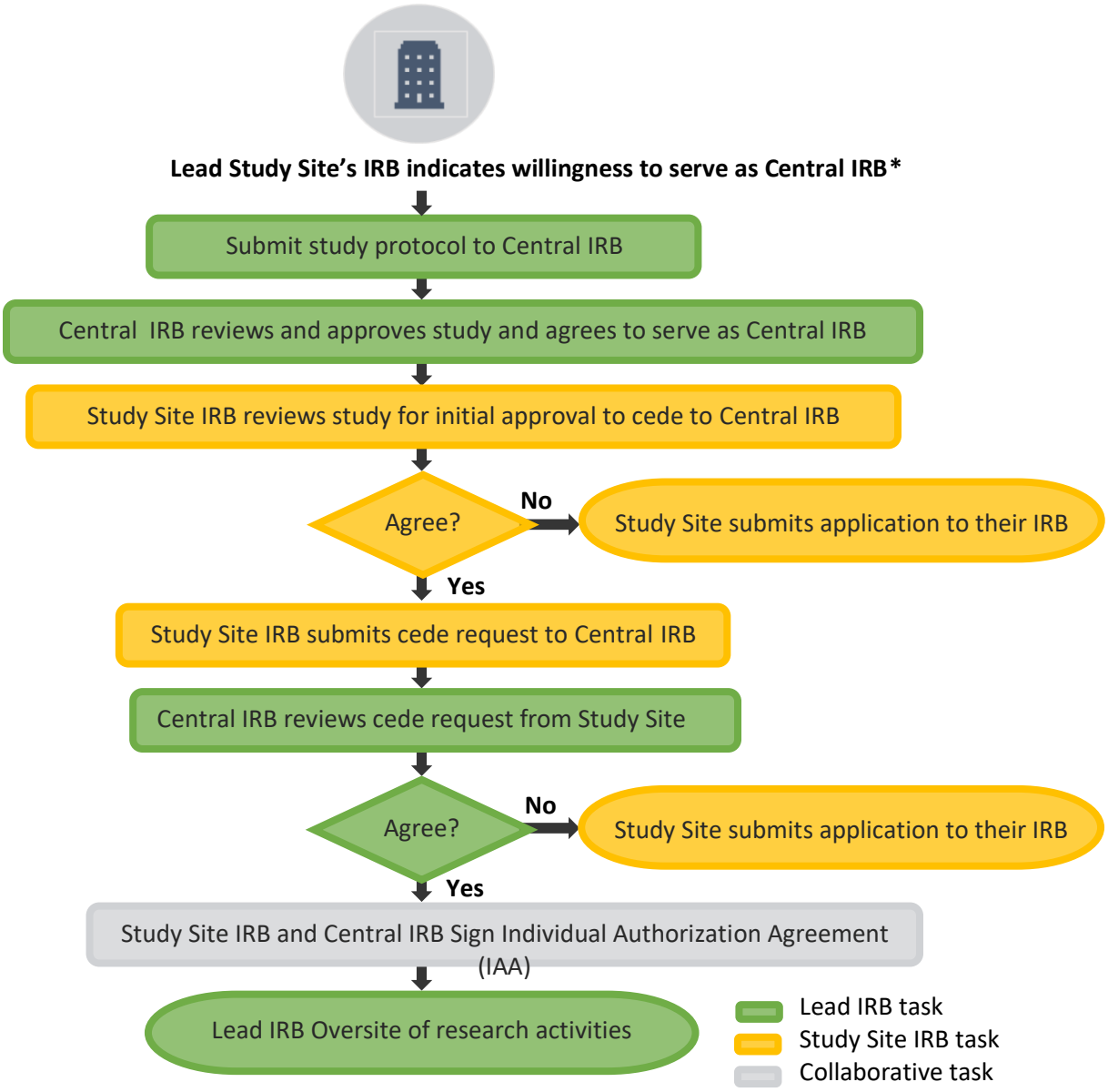
- Human Subjects Research IRB application
- Study protocol (draft protocol or proposal is also acceptable)
- Health Insurance Portability and Accountability Act (HIPAA) waiver of consent as the use or release of personal identifiable information (PII; e.g., social security number [SSN], name, date of birth) requires a written authorization from an individual or a waiver of authorization from the IRB acting as the Privacy Board.
- A Collaborative Institutional Training Initiative (CITI) or equivalent training certificate for Human Subjects Research for all study team members
- Other site-specific documents required by the Lead IRB

The pilot project utilized a Human Subjects Research application because Protected Health Information (PHI; e.g., SSN, name, date of birth) is required to link patients included in Study Site databases to NDI data. In addition, the project was ascertaining patients' vital status from NDI linkage, and there was no assurance that all individuals selected for NDI matching would be deceased (i.e., potential existed for transferring PHI information on living individuals to the NDI). Human Studies designation may vary by study and institution, and thus Coordinating Centers and Sponsors of future multi-center studies with a NDI data linkage component should follow the Office for Human Research Protections (OHRP), Human Subjects Regulations Decision Charts[8] and consult with the lead IRB of record to determine if an activity is research involving human subjects. HIPAA regulations allow the IRB to waive use of an individual authorization form criteria are met.[9]

B. Centralized IRB review process

In a centralized IRB review process, sites participating in a multi-center research study agree to rely on a single IRB of record.[10,11] We refer to the lead IRB in such multi-center studies, or the IRB of record, as the 'Central IRB' in this chapter. Utilizing a Central IRB maximizes efficiency and decreases administrative burdens to participating institutions by avoiding duplicative reviews by multiple local IRBs.[11] Furthermore, the new Common Rule requires use of a central IRB for multi-center research, with certain exceptions (see 82 Fed. Reg. at 7265 [final rule § .114]).[6] **Figure 2** provides an overview of the IRB process for multi-center research studies requiring NDI data.

Figure 2. Overview of Institutional Review Board (IRB) process for multi-center research studies requiring NDI data



** For pilot project the Coordinating Center's IRB served as the central IRB for the study, future studies could choose a different approach.*

Below is an outline of the major steps required to achieve Central IRB approval for a multi-center study.

1. Coordinating Center submits a formal request to the selected Lead IRB requesting them to serve as the Central IRB for each Study Site.
2. Selected lead IRB determines whether to take responsibility for each ceding organization individually, after reviewing risks and benefits to the study patients and site organizational policies and procedures.

3. Study Sites obtain approval from their local IRB to cede review. This includes scientific review by respective local IRB's to determine agreement to cede.
4. If selected lead IRB agrees to serve as the reviewing IRB, an IRB Authorization Agreement (IAA) between the central IRB and each study site IRB will be required. The lead IRB will initiate this agreement and is the final step in the process to activate a Central IRB. An IAA agreement will be required if no Master Common Reciprocal Agreement exists between the Central IRB and participating site. If a Master Common Reciprocal Agreement exists, they may be leveraged to improve efficiency if both lead IRB and participating study site(s) are participants in the respective research networks. This eliminates the need for re-signing legal agreements on a study-by-study basis.

It is important to discuss using a Central IRB with Study Sites immediately upon study initiation. The Coordinating Center should review and discuss any site-specific policies and procedures that need to be accommodated in the review process (e.g., state IRB reviews, technical risk assessments, and Master Reliance Agreements). Study teams should consider utilizing Master Reliance Agreements (e.g., SMART IRB[12]) where possible, especially if all study sites are members, as IAAs are already in place and assist with expediting the IRB approval process. The Coordinating Center should draft one HIPAA waiver form for the study and ask sites to review and incorporate each site's respective data disposition and security measures. More information and details regarding the Central IRB process can be found in SMART Common Reciprocal Institutional Review Board Authorization Agreement Standard Operating Procedures.[13]

The Coordinating Center should also inform Study Sites and the Central IRB that the NDI requires Study Sites to complete and sign a confidentiality agreement describing how patient-level NDI data will be protected as well as NDI required data disposition policies (i.e., data must be disposed of within five years of the NDI application submission). More details regarding this are listed under section III below.

III. Completing the NDI application

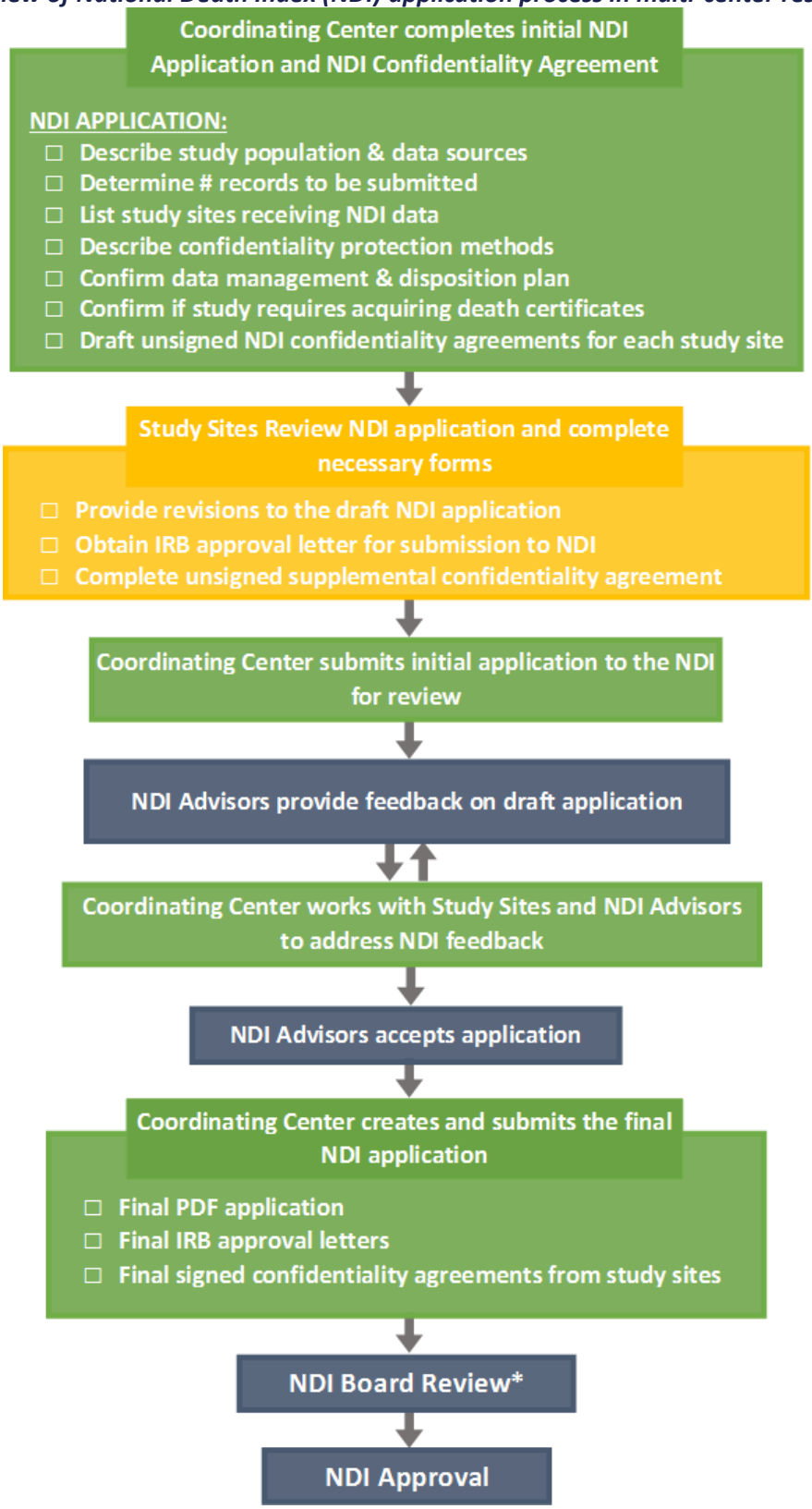
A draft NDI application (latest version available here: https://www.cdc.gov/nchs/ndi/index.htm) should be submitted to the NDI for review by NDI advisors as soon as the study population and study design are confirmed. While NDI will not accept a final application until IRB review is complete, an initial NDI review can occur as soon as a study protocol is available, and in parallel to IRB review processes outlined above. Additionally, we learned on September 23, 2020, that the NDI launched a new software application called the NDI portal. This new resource provides an electronic application process. The NDI is encouraging study teams to request access and take advantages of the streamlined process. For more information and steps in the process visit https://www.cdc.gov/nchs/ndi/portal.htm.

During the pilot study, we found the most important items to include in an NDI application were a succinct description of the study goals and purpose for linking to the NDI, data use plans, and clear statements that NDI data will not be re-released to others or reused without further NDI approvals. Data disposition and protection measures must be described in NDI required legal agreements. Designation of Sponsor, Coordinating Center, and Study Sites and a description of all entities with access to NDI data was also important. For the final application, NDI provided directions about executing legal agreements and final application submission. **Appendix A** highlights the most important aspects of the NDI application, recommended language, and feedback provided by the NDI advisors during the pilot study. Additionally, we include an example of an accepted application for NDI review board based on the pilot

study application in **Appendix A, Section VII.** The NDI also provides an updated eligibility requirements list on their website, which future research teams should consult.[14]

Figure 3. Provides an overview of the NDI application process implemented in the pilot study. For more detailed information regarding application steps, refer to **Appendix A, Section II**. After NDI advisors review the draft application and draft legal agreements, the Coordinating Center should work with each Study Site and the site and the Central IRB as needed to address any concerns raised by the NDI and attain required IRB approval letters received from each site once IAAs are executed.

Figure 3. Overview of National Death Index (NDI) application process in multi-center research settings



**NDI Board may request edits or changes; in such instances the Coordinating Center works with Study Sites to address.*

To submit a final NDI application, the following documents were required: A complete NDI application document; executed legal Confidentiality/Supplemental Confidentiality Agreements from each Study Site, the Sponsor, and the Coordinating Center; and an IRB letter from each Study Site. We learned during our application process that it is standard NDI procedure to provide electronic and original paper copies of all elements of the NDI application, including all executed legal agreements were required to submit to the NDI for review by the NDI Advisory Board. As policies may evolve over time, we advise future studies check with the NDI to confirm what is required for the final NDI application submission. The NDI Advisory Board is comprised of a multi-member panel that reviews and determines whether to approve an NDI application.

The NDI Advisory Board review process and timeframe varies depending on the number of requests for the NDI data and advise budgeting at least four to six months for this process. We advise working closely with NDI staff to complete the application and to ensure the proper steps are followed to assist with NDI Board review. Following NDI Board review, the NDI will email the Lead Principal Investigator (PI) at the Coordinating Center with the determination and will provide guidance regarding any adjustments needed for the NDI application to finalize the approval. This may require working with Study Sites to adjust the application or confidentiality agreements as needed. Following final edits, the NDI will confirm approval and send an email to the Lead PI and provide an approval letter indicating the type of search approved (i.e., routine NDI search, NDI plus with vital status known or unknown) and will provide the application number as well as information regarding how to prepare records for submission to the NDI. For more information regarding preparation of files, refer to **Chapter 2**. The study application number must be used in all communications with the NDI.

The latest NDI approval criteria are listed on the NDI website (latest version available here: https://www.cdc.gov/nchs/data/ndi/ndi_approval_criteria.pdf), along with more information about the NDI's data disposition policy. The NDI requires that in most circumstances that NDI data must be disposed of within five years of NDI application submission. Any applicant that does not plan to follow this NDI policy must submit a justification of why the identifiable data needs to be maintained beyond five years of receipt of the initial application or beyond a previously approved extension period. More details regarding specific NDI requirements are listed on the NDI website and in the *NDI User's Guide* and the NDI Data Disposition Form is provided **Appendix A, Section VII.** This is important to plan for during the initial application process, and individual institutions must be sure to implement procedures to follow these policies. In the pilot study, we found Study Sites had different institutional data disposition policies, and the Coordinating Center worked with each site to ensure sites were able to adhere to the standard NDI data disposition policy.

It important to note that the NDI charges fees for routine NDI searches that consist of a service charge plus nominal fee per user record for each year of death searched and the NDI plus search is slightly higher. Furthermore, there is a fixed fee per decedent if individuals are all known to be decreased. For more pricing information, refer to the **NDI User Fees** document for all fees and for a worksheet to assist in calculating your total charges for an NDI search. Discounts are also provided for large volumes (more than 100,000 records) for an NDI search. We also learned that on January 14, 2020, through an agreement between the National Institutes of Health (NIH) and NCHS, the NIH will reimburse the NCHS NDI for the costs of NIH-supported investigators to link their research databases with the NDI for the research aims supported by the NIH.[15] This may be helpful for future studies funded by the NIH to consider in their budgets. In addition, the NDI will also provide an NDI Fee Worksheet and NDI Transmittal Form upon approval of the NDI application that must be completed for payment of records submitted to the NDI. For more information regarding these forms, refer to **Chapter 2, Section V.**

Following NDI approval, the Coordinating Center informs all Study Sites of the approval and provides the final approved application, application number, and NDI fee worksheet and transmittal forms that will be completed when records are ready for submission to the NDI. An example of the NDI fee worksheet and transmittal form is provided in **Appendix B (Section III and IV)**.

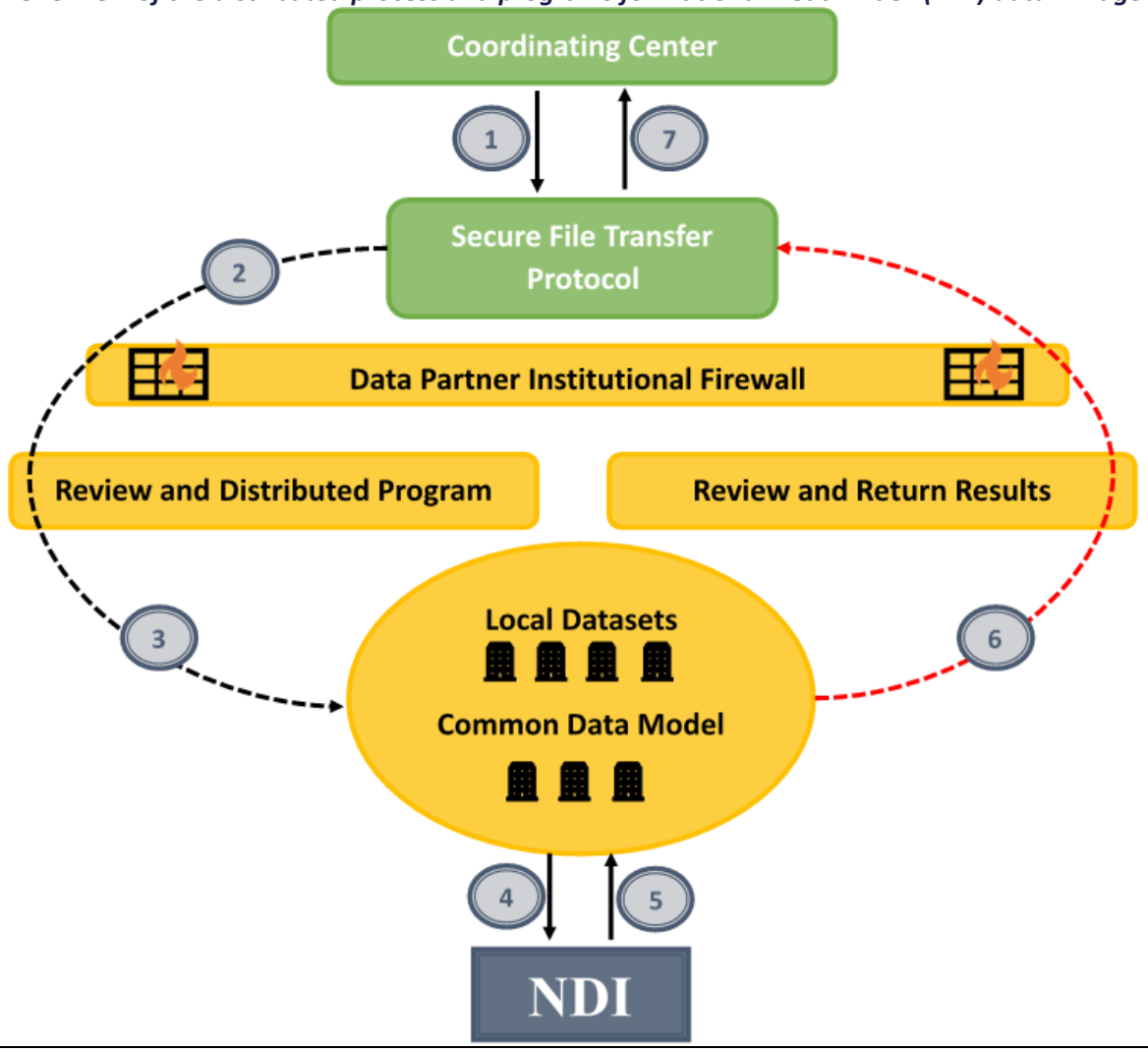
Chapter 2: Preparation and Submission of Files to the National Death Index for Matching

This chapter describes a distributed process for preparing and submitting files to the NDI in multi-center studies. This standardized and reusable distributed technical process was designed to efficiently attain and analyze death and cause of death information from the NDI across multiple Study Sites without sharing PHI between health plans or with the Coordinating Center. Preparing and submitting files to the NDI for matching is just one piece of the distributed multi-step linkage process (**Figure 4, steps 1-4).** **Chapters 3-5** describe additional technical steps required to attain and analyze death and cause of death information from the NDI in a distributed fashion. Please note, although we used SAS 9.4 ® software[16] to create these programming packages the methods we developed and described could be implemented with other software. During the pilot project, we chose to beta-test versions of each program prior to distributing a final version of the program to all Study Sites. This approach ensured that each program performed as expected, uniformly across sites, and avoided multiple re-runs of the final program by all Study Sites. To avoid duplicative program runs, future studies may also wish to consider using this approach.

Table 2. Programs in the distributed National Death Index (NDI) linkage process

Distributed Program Number	Program Description	Chapter
Distributed Program 1	Identify the cohorts of interest	**Chapter 2, Figure 5**
Distributed Program 2	Select which patients to submit to the NDI for matching	**Chapter 2, Figure 6**
Distributed Program 3	Quality check files prepared for submission to the NDI	**Chapter 2, Figure 9**
Distributed Program 4	Create outbound files in the NDI required format,	**Chapter 2, Figure 10**
Distributed Program 5	Process and check returned NDI data, and determine approach for saving the 'best' match from all potential matches	**Chapter 3-4, Figure 13**
Distributed Program 6	Save the best NDI match, and create files to be used in final study analyses	**Chapter 5, Table B3., Table B5.**

Figure 4. Overview of the distributed process and programs for National Death Index (NDI) data linkage process

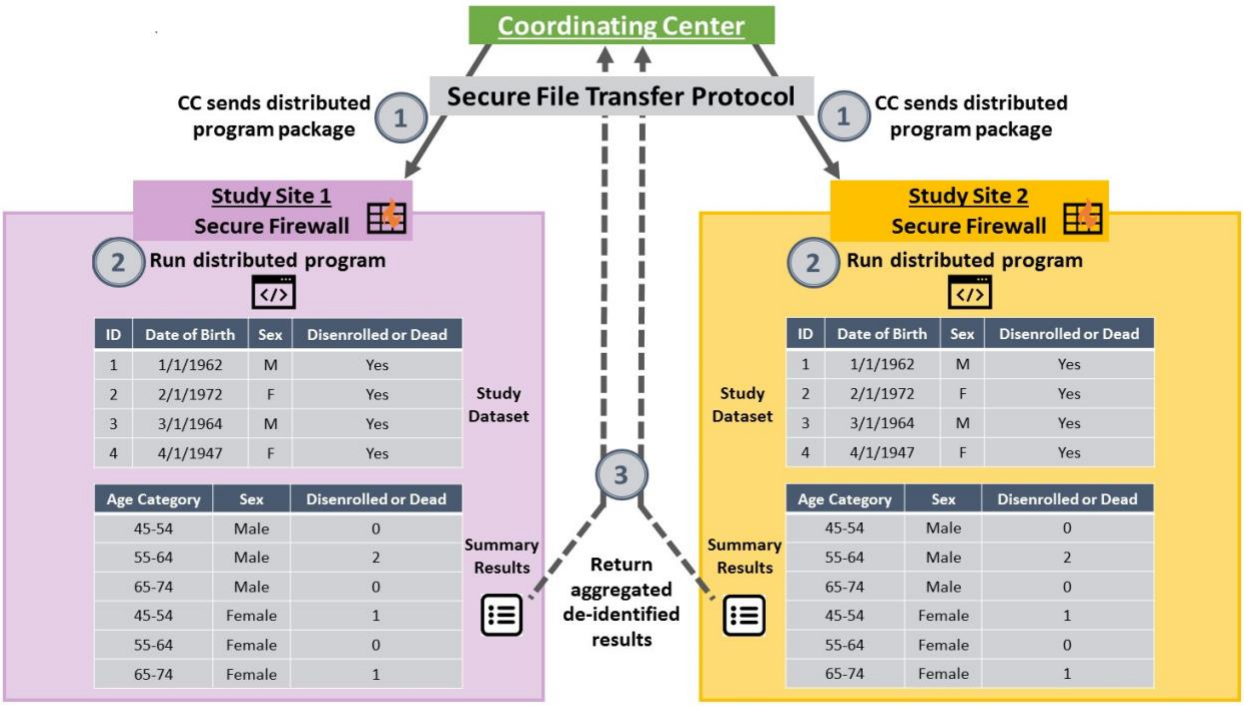


1. Coordinating Center initiates a distributed program identifying patients of interest
2. Participating Study Sites retrieve the distributed program via a secure and HIPAA compliant secure transfer protocol system (i.e., SSH File Transfer Protocol)
3. Participating Study Sites review and run the distributed program identifying patients to submit to NDI
4. Participating Study Sites prepare files to submit to NDI
5. NDI returns files to participating health plans
6. Participating Study Sites run distributed program developed by the study team against returned NDI+ files to identify matches to be saved
7. Participating Study Sites remove all protected health information, save other data to project analytical files, and return results for review by the study team

I. Identifying study population of interest

The first step to conducting an efficient distributed linkage of multiple Study Site databases to NDI data is identifying cohorts of interest at each of the Study Sites. During the pilot study, we identified use case cohorts using the publicly available Sentinel Cohort Identification and Descriptive Analysis (CIDA) program, version 7.0.0[17] in combination with additional code to provide descriptive information to the study team. This combination of the CIDA program and additional code is the first program package in the distributed NDI data linkage process, and the general approach can be modified and reused by future study teams. In our pilot project, we submitted patients with deaths recorded at the study sites, as well as those with potential deaths to the NDI for linkage. We defined potential death as health plan disenrollment between cohort entry and cohort exit plus 365 days, without subsequent re-enrollment or medical or drug utilization >60 days post-disenrollment, but this could be modified by future studies.[4] For more information about the piloted use case specifications, please refer to **Appendix B, Section I**. Study Sites retrieve this distributed program, developed by the Coordinating Center, using secure methods (**Figure 5**).

Figure 5. Distributed Program 1: Coordinating Center (CC) sends programs to identify cohorts, and flag cohort members to send to the National Death Index (NDI) for linkage



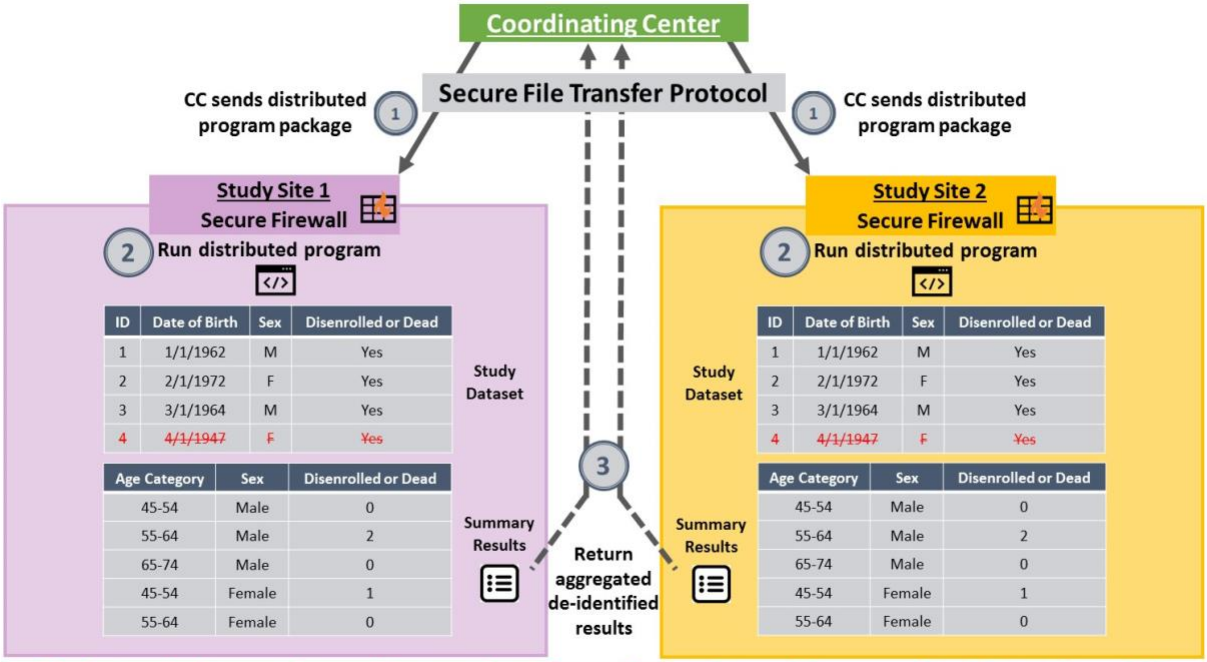
II. Selecting patients to submit to the NDI for matching

Each study will need to determine exactly which patients to submit to the NDI for linkage and may also consider whether or not to submit all study patients to the NDI for linkage or just study patients whose vital status is unknown. While in some instances it might be preferable to submit all patients captured within a particular study to the NDI for linkage, it is possible that the number of patients eligible for the NDI data linkage may be larger than study resources allow. To address this potential, we built options

into the developed NDI linkage process to allow future studies to manage the overall number of records being submitted to the NDI for matching. During our pilot study, we submitted patients with deaths recorded at the study sites, as well as those with potential deaths to the NDI for linkage. Please see **Supplemental Appendix** for technical specifications.

Distributed Program Package 1 provides aggregated information about the study patients that can be used to examine baseline characteristics of the study population and determine which patients to submit to the NDI for linkage. This program also saves patient-level files for the study population at each Study Site, which are processed in subsequent steps of this NDI linkage process. The study team can use filtering flags created by the program to manage the overall number of records submitted to the NDI for matching.

Figure 6. Distributed Program 2: Coordinating Center(CC) sends programs for applying filtering criteria to cohorts for volume control and to finalize submissions to the National Death Index (NDI)



Features of the distributed program which are designed to be reused by future studies include the following tables and variables:

A. Baseline table

Distributed Program 1 provides an aggregated table describing the identified cohorts of interest for review by the study team. Variables and code lists to identify these variables could be modified as needed for use by future studies (see **Appendix B, Section II**) for example based on the piloted use case).

B. General filtering flags

We developed a flexible option which outputs a standardized aggregated table with additional information allowing for management of the overall number of records being submitted to the NDI for matching. This allows the study team to apply filtering criteria to the cohorts should it be important to

refine cohorts or reduce sample sizes. This table contains counts for the number of patients by the following strata:

- Study Site
- Cohort (exposed, comparator)
- Drug Name (name of index drug exposure, used for cohorts defined by multiple drug exposures)
- Age Group (flexible age categories defined by study team)
- Sex (male, female)
- Index Year (study years, e.g., 2000-2017)
- Drug Therapy Length (days supplied in weeks)
- Reason Selected for NDI matching (potential death, disenrollment, both, or not selected)
- Potential Death in Study Site data (source of death information, alive/no evidence of potential death)
- Interval of Potential Death, if any (alive, no potential death occurred, potential death occurred from episode index date through episode end date, potential death occurred from episode end date + 365 days, potential death occurred within some other time interval)
- Final Enrollment Status (disenrolled during episode or 365 days following without reenrollment, or enrolled)
- Time from Episode End to Disenrollment (grouped in 30-day increments)
- Medical or Pharmacy Utilization observed within the 60 days after date of last disenrollment (yes or no)
- Minimum Necessary Enrollment Observation Period (captures whether minimum necessary observation time defined by study team was available at the Study Site; the observation period in the pilot project was cohort exit date + 365 days)

Counts based on stratifications of all variables listed above can be provided to the study team. These strata and counts enable the study team to refine the numbers of individuals to be submitted to the NDI for matching as needed. If resource constraints or other considerations limit the number of cohort members that can be submitted to the NDI for matching, the study team can choose to make selections from among the strata variables to achieve a lower number of cohort members for submission (e.g., restricting submission to specific age groups or study years).

C. Cohort specific filtering flags

In multi-center studies, it is possible that the study team may wish to exclude cohort members from matching to NDI data based on diagnoses or additional drug exposures. For example, in the pilot study use case the study team wished to consider excluding patients with malaria diagnoses during the baseline period because these patients may be prescribed an antiarrhythmic medication for a different indication (e.g., quinidine). In addition, the study team might wish to consider excluding patients taking a specific medication during the baseline period. For these reasons, we created two optional cohort specific filtering flags that can be used to filter cohorts:

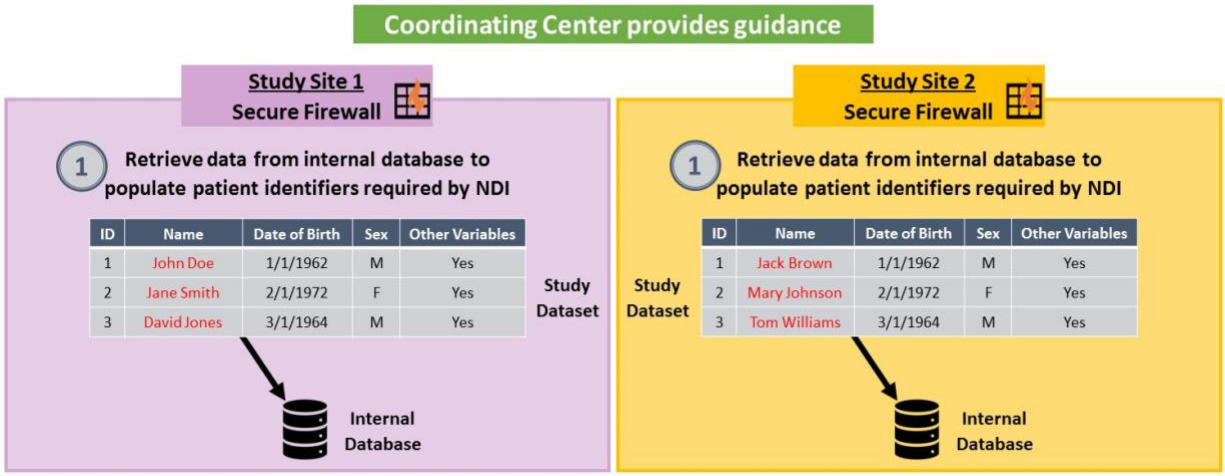
- Baseline condition flag (optional, pre-specified with diagnoses or procedure codes by the study team)
- Baseline drug exposure drug flag (optional, pre-specified with National Drug Codes [NDCs] by the study team) if the outcome or drug flag option is selected and specified by the study team, counts of cohort members with the baseline condition or drug exposure will be returned for review by the study team and can be used in addition to the general filtering flags to refine the numbers of individuals to be submitted to the NDI for matching.

After the study team reviews the baseline table and filtering criteria, they need to choose whether to apply any cohort filtering logic. The Coordinating Center then modifies the program according to the needs of the study team by applying any additional filtering desired and submits it to the Study Sites for execution. Study Sites execute the program that applies cohort filtering logic to finalize the list of cohort members that will be submitted to the NDI for matching. This is the second program package that Study Sites run in the NDI distributed linkage process (**Figure 6**).

III. Preparing files for submission to the NDI

A. Attaining data elements required by the NDI for linking

Figure 7. Study Sites populate data elements that are required by the National Death Index (NDI) for matching (no distributed program)



NDI publishes patient identifier information that Study Sites must provide in order to conduct a NDI data search, as well as required file structures in their *NDI User's Guide*.[18] To be eligible for an NDI search, each record must contain at least one of the required combinations included in **Figure 8**. The NDI will also accept additional variables and suggests providing as much information as possible to improve matching quality.[18] Study Sites have protections on the PII especially SSN and require multiple levels of privacy, compliance and IT security reviews. A brief summary of scientific rationale for the need of SSN was required by one Study Site and was useful in attaining approval. An example of this summary is provided in **Appendix B, Section V** and can be adapted for individual Study Site needs.

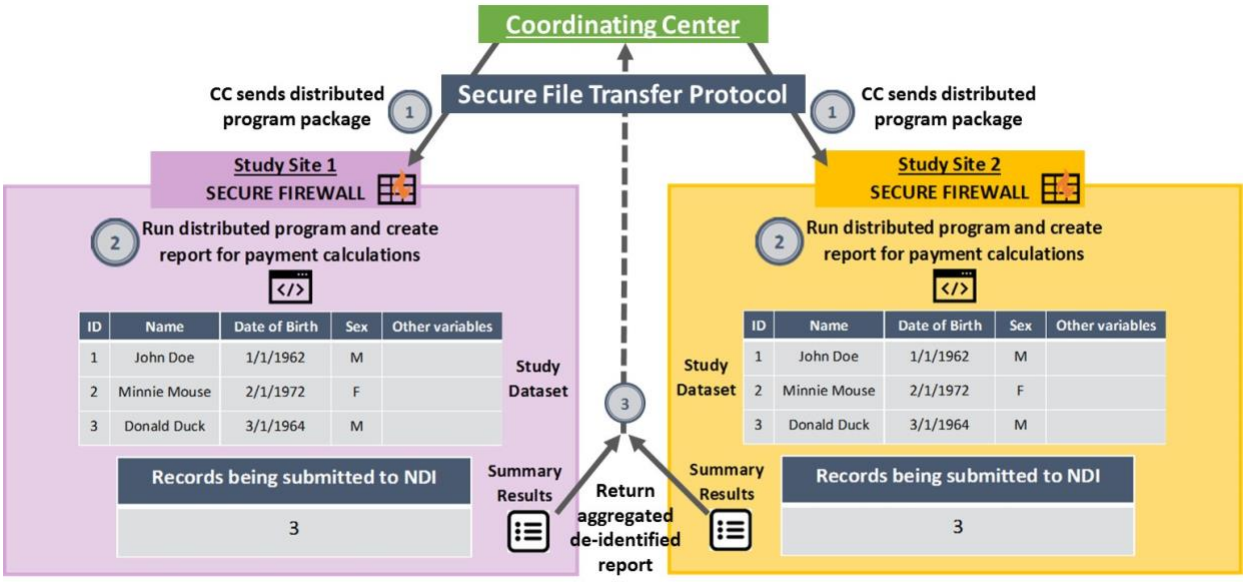
Figure 8: matching variables used by the NDI

All variables Accepted by the NDI	***Required variable Combinations***
1. First Name 2. Middle Name 3. Last Name 4. Date of Birth 5. Sex 6. Father's surname 7. Social Security Number (SSN) 8. Marital Status 9. State of last known residence 10. State of birth 11. Age at death (if known) 12. State of death 13. Date or year of death (if known) 14. Date or year of last contact	1. First AND last name AND month AND year of birth OR 2. First AND last name AND SSN OR 3. SSN AND month AND day AND year of birth, AND sex

Study Sites will access required data elements from their source systems (e.g., claims and enrollment systems) and insert these data elements into data files structured by the Study Team, for transmission to the NDI for matching. There are no distributed program packages for Study Sites to run in this step, as the Study Sites must manage the process of populating identifiers needed by the NDI (i.e., required identifiers may not be available in the main study database, and may need to be located elsewhere). .

B. Ensuring study subject records are eligible for an NDI search

Figure 9. Distributed Program 3: Coordinating Center sends program to Study Sites designed to conduct quality assurance of records prepared for NDI submission



To ensure files meet NDI's requirements, the Coordinating Center will distribute a program for local execution by the Study Sites to identify any potential data or formatting issues. The checks focus on data

quality. For example, given the importance of SSNs in NDI matching, there is a series of checks designed to indicate an invalid SSN:[19]

1. Invalid whole number: Some specific 9-digit SSNs are considered invalid. These are as follows:
   a. All 9 positions with the same digit; e.g., 999999999
   b. Specific voided numbers: 078051120[20]
   c. Specific range: Any in the range of 987-65-4320 through 987-65-4329
2. Invalid values in positions 1-3 (area numbers): 000, 666 and any of 900-999.
3. Invalid values in positions 4-5 (group numbers): 00.
4. Invalid values in positions 6-9 (serial numbers): 0000.

Other examples of checks are the following: names should contain only letters and allowable punctuation (e.g., hyphen or apostrophe), sex should be only "M" or "F", etc.

To ensure adequate quality for all study sites, Study Sites will be required to execute the program (**Figure 6**), which checks populated data elements against NDI criteria for acceptance. This produces:

a. A Study Site patient-level report of data quality issues
b. A Coordinating Center aggregate-level report with counts of quality control issues

An example of a fictitious Study Site patient-level report is included in **Table 3**.

Table 3. Example of fictitious Study Site patient level report, remains at each individual Study Site

Patient ID	Variable ID	Variable Description	Specific Details
A123	SSN_1	SSN contains values other than digits	09852445A
B456	SSN_3	Invalid SSN	999999999
C789	Day_of_Birth_2	Day_of_Birth contains an invalid value for Month_of_Birth	Month_of_Birth=4, Day_of_Birth=31

All patient identifiers remain at the individual Study Sites, and the Coordinating Center will receive only an aggregate report describing only the type and magnitude of quality issues at each Study Site. No patient identifiers will be shared between Study Sites, or with the Coordinating Center.

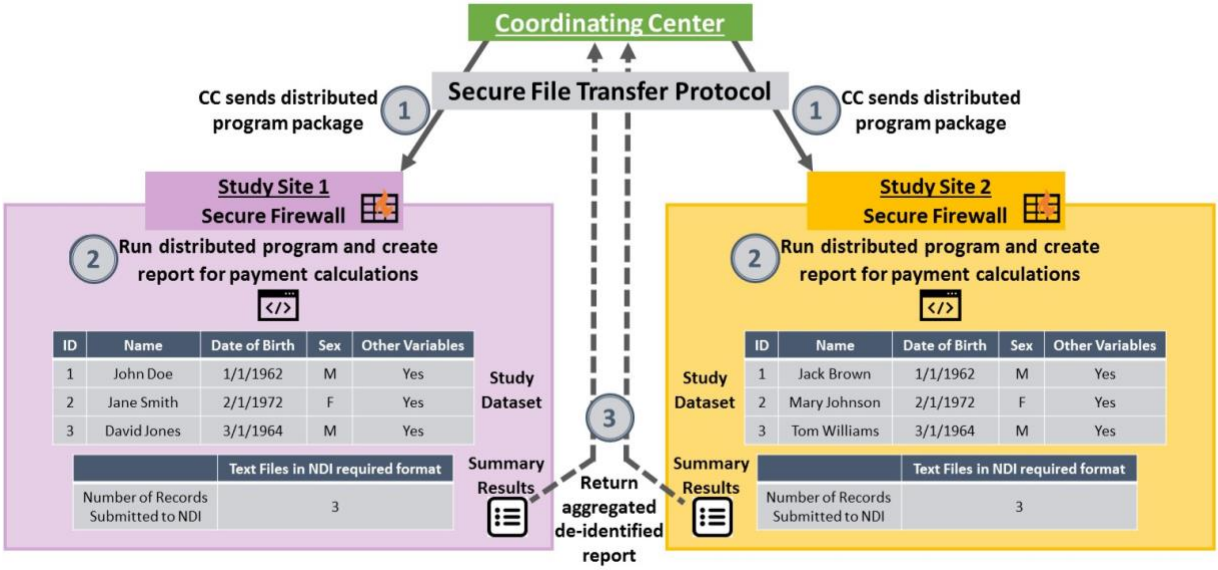
If a patient has multiple names, or multiple last names, the NDI allows for creation of what they term "alias" records,[18] which simply means the NDI will conduct their search also utilizing each individual "alias" name. For example, the NDI recommends that hyphenated last names be also submitted as distinct alias records. With a patient last name of "Dexter-Murphy," there would be three records with the last name field of these three records being respectively: Dexter-Murphy, Dexter, and Murphy. The latter two records would be considered the "alias" records. This feature is intended to improve matching rates and quality of matches. Thus, the developed quality checking program automates the creation of these alias records according to NDI specifications in situations where patients have multiple names.

The execution of the quality assurance program may need to occur multiple times until Study Sites correct their data or remove patient records that cannot be fixed. An example is when the information from the Study Site source has been confirmed, but is still invalid, such as an invalid social security number that has been maintained by a health insurer's enrollment system. When the results of the quality assurance program indicate that all records are ready for sending to the NDI, then this program will have achieved its goal of ensuring that all records sent by all Study Sites are acceptable by the NDI for linking.

The developed quality assurance program can be adapted and reused for future multi-center research studies. This is the third program package that the Study Sites run in the NDI distributed linkage process. Lessons learned during file preparation and quality control processes that would be useful to future studies are documented in **Appendix B**.

IV. Creating and submitting files in the NDI required format

Figure 10. Distributed Program 4: Coordinating Center (CC) sends programs to Study Sites designed to export quality-checked patient level data to the National Death Index (NDI) required format

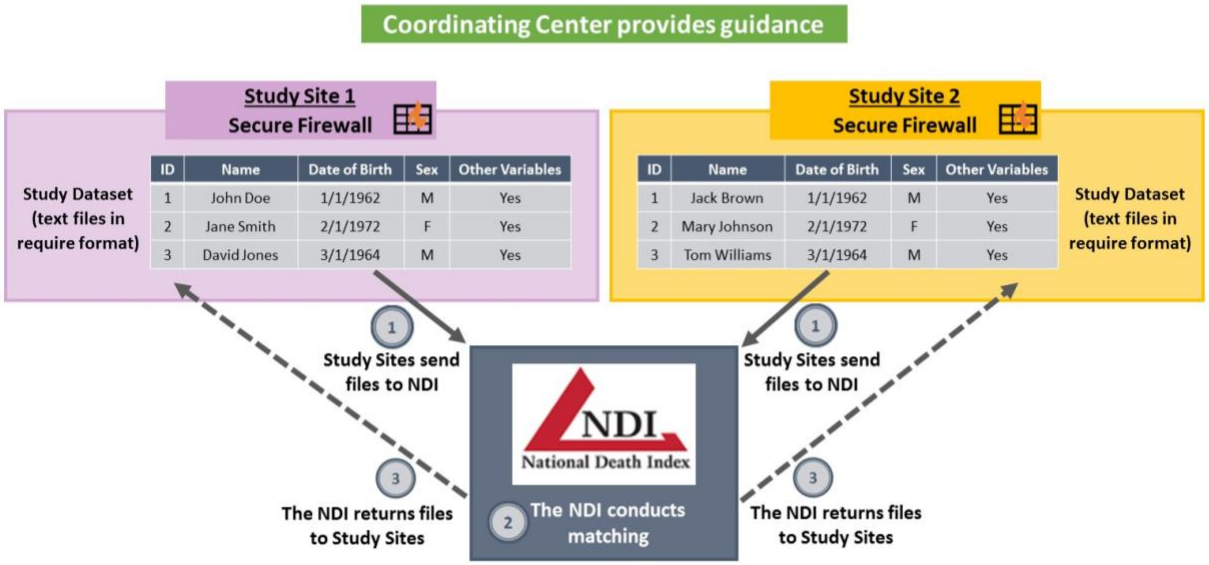


After files intended for submission to the NDI have been checked to ensure sufficient completeness and quality, each Study Site will execute the fourth distributed program package, which will export the quality checked patient level data described in Steps 3-4 to flat text files for submission directly to the NDI for matching (**Figure 10**). This program will also generate aggregate reports of records being submitted, so that the NDI Transmittal Form and the NDI User Fees worksheet can both be accurately completed, and payment calculations can be accurately performed.

Study Site data files will be transferred to the NDI via either password-protected encrypted data CDs or SSH secure file transfer protocol (SFTP), according to Study Site and NDI requirements, and NDI will conduct their matching activities (**Figure 11**). It is important to note that the *NDI User's Guide* does not describe the file transfer methods that the NDI will accept (CD-ROM or SFTP). We learned that the NDI prefers CD-ROM submissions and requires permission from the NDI to use SFTP. If study site would like to use SFTP, it is important to reach out to the study contact at the NDI to coordinate SFTP site access, and whether sites will be using NDI's SFTP site or their own SFTP site. For sites that choose to use NDI's SFTP, the latest version of WinSCP[21] must be used and recommended to connect with the NDI to confirm the acceptable version. In addition, it is recommended to setup a test for data exchange to ensure both parties can access data files. Determining current acceptable methods in advance of file submission will save Study Sites time, especially in multi-center studies.

More details about the NDI's matching procedures and data exchange are included in **Chapters 3** and **4**.

Figure 11. Study Sites send files to the NDI, and the NDI conducts matching



V. Completing and submitting the NDI pricing sheet

The NDI requires all studies to submit the NDI User Fees worksheet (See **Appendix B, Section III**), summarizing the number of records that will be submitted and the number of years that need to be searched; this information is used to calculate user fees. As noted above in Step 5, each Study Site will have, from the results of the fourth distributed program, the number of records by specific years that need to be searched. Payment needs to be received before NDI will conduct the linkage and this will be managed and coordinated by the Coordinating Center. Study Sites must submit a separate ***NDI Transmittal Form*** for each file and a ***Worksheet for Calculating NDI Charges*** with study file.[18] We recommend that the Coordinating Center provide payment for the NDI charges to cover the NDI fees for the project. The Coordinating Center should confirm payment information once final record charges are confirmed from all Study Sites to include on the NDI Transmittal Form. The Study Coordinating Center also worked with Study Sites to verify payment amount, that included an initial service charge for the project, charged to one Study Site, and reduced service charge for each subsequent site submission. Additionally, after charges are verified by the Coordinating Center, the NDI requires that the NDI Transmittal Form is signed from a person authorized to request the NDI search. For Study Sites, the Site Principal Investigator signed these forms.

Generally, in the first six months of every year, the NDI provides early release files which contain data that have not yet been finalized for the previous year. When the early release data are available, the NDI notifies the Study Site, and offers to conduct the same search of the data against that final year at no charge. with a nominal service fee per request. During the pilot project we learned it was important to consider if early release request would be included in the study, as the NDI requires early release[22] request records should be broken out and submitted as a separate file, rather than one single file containing all request data. A separate NDI Transmittal Form and Worksheet also must be completed for calculating early release charges. The NDI maintains information about the status of their early release files on their website (see here for more information: https://www.cdc.gov/nchs/ndi/ndi_early_release.htm).

Chapter 3: Receiving Files from the NDI

I. Conducting record linkage at the NDI and sending files to study sites

When the NDI receives data files from the Study Sites, their technical staff will attempt to link submitted study patients to death and cause of death records. In our distributed NDI linkage process, each Study Site exchanged files individually with the NDI. The NDI will process each Study Site's file sequentially, and thus each Study Site can submit data files to the NDI as soon as their data files have been prepared and are deemed of sufficient quality (See **Chapter 2** for details).

The NDI follows specific data linkage processes documented in the *NDI User's Guide*.[18] The linkage process is entirely under the control of the NDI and submitters cannot manage the linking algorithms or process. When the NDI completes data linkage activities, the NDI creates data files containing potential matches for each individual user record submitted to the NDI by the Study Site. Please refer to the *NDI User's Guide*[18] and **Chapter 4** for information and variables the NDI provides to Study Sites. We also include detailed lessons learned during our pilot study that could be useful for future studies in **Appendix B**.

II. Receiving and importing the NDI files

The NDI will return data files directly to the Study Sites through the same secure methods mentioned above (i.e., encrypted CDs or secure file transfer protocol). These NDI data files will remain behind the Study Site firewalls and will not be shared with the Coordinating Center or other Study Sites.

We developed a distributed program (Program 4) which serves multiple functions including: A) importing the returned NDI data into datasets and preparing them for use in future steps of the distributed technical process, B) comparing returned NDI files against submitted NDI files to ensure that all submitted records from health plans to the NDI are found in one of the three returned NDI files, C) characterizing the number of potential NDI matches for each submitted patient record and implementing various algorithms for selecting the 'best' match from all returned NDI matches (discussed further in **Chapter 4**), D) providing summarized aggregate data files which Study Sites can share with the Coordinating Center to ensure standardized processes are followed across sites.

As with previous distributed programs, the Study Sites run the program to perform these functions, and only returns aggregated counts to the Coordinating Center, thereby maintaining the distributed process. **Chapter 4** contains more details about this fifth distributed program that Study Sites run in the NDI distributed linkage process, and we include lessons learned that would be useful to future studies in the **Appendix B** and technical specifications for **Distributed Program 5** are provided in an Supplemental Appendix. An overview of the data NDI returns, and considerations for use are also included in **Chapter 4**.

Chapter 4: Selecting the Best Match from Data Returned by the NDI

When the NDI completes data linkage activities, the NDI returns files containing single or multiple potential matches for each individual user record submitted to the NDI by the Study Sites. The NDI states selecting the best match from these potential matching death records or rejecting poor matches is the responsibility of the end user. Specifically, NDI records involved in matches should only be considered 'possible' matches, and the end user must determine which matches are 'true' matches, which are 'false', and which are 'questionable' and require further investigation.[18] *The NDI User's Guide* further describes procedures that NDI uses to conduct matching activities.[18]

In this chapter, we provide an overview of information the NDI provides to assist with selection of the best match and describe a distributed technical approach for selecting and retaining the best match from files that NDI returns to Study Sites. We designed a reusable and standardized process to select the best NDI match that is flexible enough to allow future studies to tailor their approach to specific study needs. In addition, we prioritized developing an automated process for selecting the best match from returned NDI data over processes that require large amounts of manual review and adjudication. We made this choice as multi-site studies often include large study populations, and manual review and adjudication of multiple potential NDI matches is often not practical. Thus, we implemented approaches for selecting the best match without manual review as described in the *NDI User's Guide*. We also reviewed automated approaches other studies have utilized for selecting the best NDI match that could be implemented in a distributed network. More details about returned NDI variables and the developed distributed technical process for selecting and retaining the best NDI match are described below.

I. Selecting the best match: an overview of variables provided by the NDI

The *NDI User's Guide*[1] provides guidelines for selection and retention of NDI matches, as multiple possible matches for each individual submitted may be provided within NDI-returned data files. This requires researchers to assess the quality of each possible NDI record match listed and to determine which possible matches are "best" matches. The NDI recommends a multi-step process when determining the best match among possible multiple matches, including utilizing the NDI provided probabilistic matching scores to distinguish true matches from false matches.

The NDI returns several variables to assist with evaluating potential matches and selecting the best match, and provides extensive information describing all variables within the *NDI Users Guide*.[1] Below we focus on the key NDI variables which are most useful in automating the match selection process and are utilized in our distributed technical approach for selecting and retaining the best match from NDI data files.

The NDI returns indicators of agreement between variables on the submitted patient record and the matched NDI record (e.g., SSN agree or disagree), as well as an exact match indicator. An exact match indicates all items provided on the user record match exactly with the corresponding items on the NDI record. In addition to these deterministic matching indicators, the NDI provides probabilistic matching variables designed to assist the study team with selecting the best match. The NDI names these variables as status code, class code, and probabilistic matching score. We developed a distributed process that utilizes NDI variables to automate the selection of the single best match from NDI data where possible and prioritizes reducing or eliminating manual record review to reconcile the final match selection.

Probabilistic matching score: Each returned NDI potential match record is assigned a probabilistic score which reflects the agreement between information on the submitted patient record and the NDI record. The probabilistic matching score itself is the sum of weights assigned to each identifying data item used

in the NDI record matching process. Weights are calculated as the base 2 logarithm of the inverse of the probability of the occurrence of the characteristic within nationally representative samples available to the NDI.[18] Weights are either positive or negative. If agreement exists between the patient record and the NDI record for a specific variable (e.g., SSN), the weight is positive. If there is no agreement, the weight is negative.[18,23]

Figure 12. NDI User's Guide, Probabilistic Matching Score Equation

$$Score = W_{SSN} + W_{firstname\ x\ sex\ x\ birthyear} + W_{middleinitial\ x\ sex} + W_{lastname} + W_{race} + W_{sex} + W_{maritalstatus\ x\ sex\ x\ age} + W_{birthday} + W_{birthmonth} + W_{birthyear} + W_{stateofbirth} + W_{stateof\ residence}$$

Abbreviations: SSN, Social Security Number; W, weight

Class code: While the probabilistic matching score reflects the weighted probability that the submitted patient record and the NDI record match (e.g., name, SSN), class codes are assigned independently of probabilistic matching score and reflect that some NDI matching variables are more important for determining high quality matches than others. Given the importance of SSN as an identifier, each NDI record is classified as to whether SSN was present and agreed (Class code 1 or 2), present but disagreed (Class code 5), or was missing (Class code 3 or 4). While records assigned Class code 1 or 2 often will have large probabilistic matching scores as matching SSNs are weighted heavily within the probabilistic matching score equation, some records with an assignment of Class code 3 or 4 may also have a large probabilistic matching scores.[18,23] For example, when a patient's name is extremely uncommon and matches an NDI record this variable may be heavily weighted, which could potentially contribute to a larger probabilistic matching score even when SSN was not submitted to the NDI for matching. Each potential NDI match is assigned one of five classes, based on which specific patient record and NDI record variables matched. Please see **Table 4** for additional details extracted from the *NDI User's Guide*.[18]

Table 4. National Death Index User's Guide, Class code definitions[18]

Class code	Definition of Class
Class 1	Exact match on SSN (or at least eight digits), first name, middle initial, last name, sex, state of birth, birth month, and birth year.
Class 2	SSN matches on at least seven digits, and one or more of the other items from Class 1 may not match. (Note: Some matched cases are moved from Class 2 to Class 5 because of an indication that the reported SSN belongs to the spouse. This includes those cases for which the SSN is known and matches, but the first name and sex do not agree.)
Class 3	SSN unknown but eight or more of first name, middle initial, last name, father's surname (for females), birthday, birth month, birth year, sex, race, marital status, or state of birth match.
Class 4	Same as Class 3 but fewer than eight items match.
Class 5	SSN is known but does not match. [Note: Some matched cases are moved from Class 5 to Class 3 because of an indication that one of the SSNs (on the user record or on the death certificate) may have been reported incorrectly, but a significant number of other data items are in agreement.]
Abbreviations: SSN, Social Security Number	

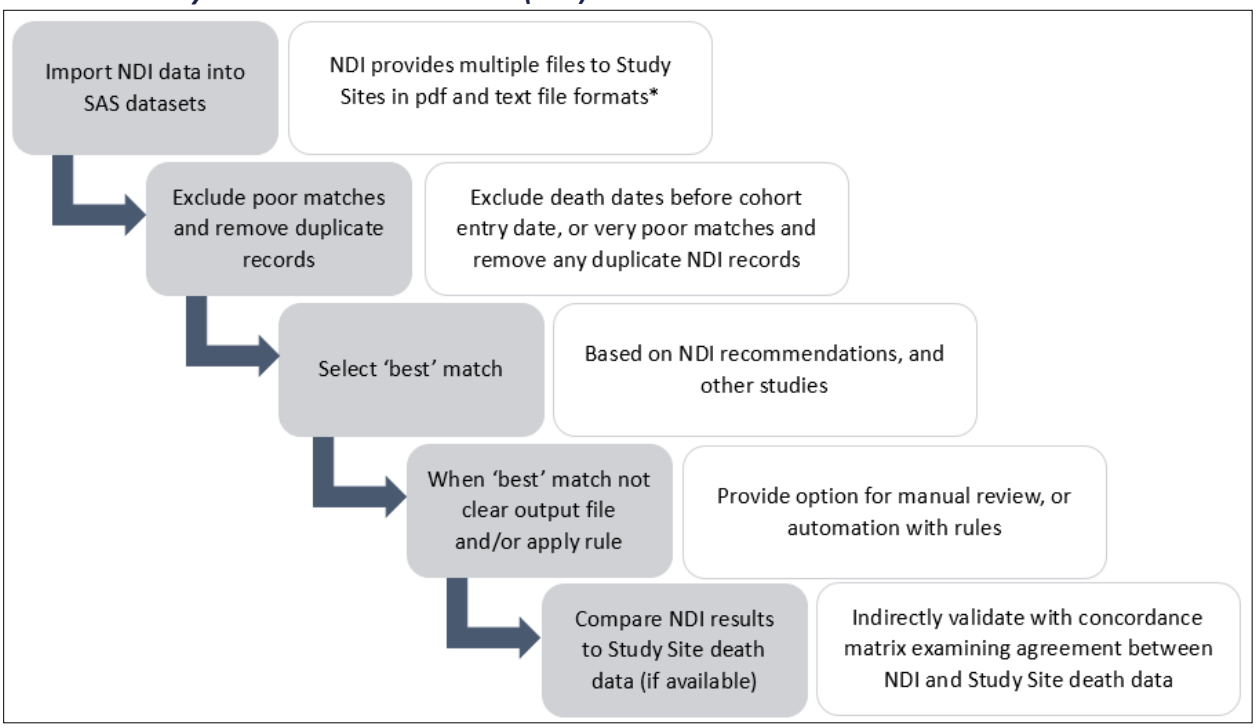
Status code: Status code is the death status that the NDI suggests. Records with a status code value of 1 are considered true matches and assumed dead; records with a status code value of 0 are considered false matches and are assumed alive. However, the NDI also states that the status code should only be used as a guide. Some potential NDI matches with a status code of 0 may be true matches, some potential matches with status code of 1 may not be true matches.[18]

II. Developing a process for selecting the best match from data returned by the NDI

While the *NDI User's Guide*[18]provides information about NDI matching processes, it states the responsibility for selecting the best match from potential matching records or rejecting poor matches rests upon the end user. In a distributed multi-site research setting, Study Sites maintain physical and operational control over their electronic health data in their existing environments.[1–3,24–26] We focused on developing a standardized reusable distributed NDI linkage process because a distributed data linkage approach promotes data sharing by protecting patient privacy, data security, as well as any proprietary interests. However, selecting the best match from data returned by the NDI in a distributed network also may require development of technical programs and processes to promote standardization and efficiency across multiple Study Sites. The distributed process we developed for selecting the best match from data returned by the NDI was informed by the *NDI User's Guide*,[18,23] the NIH Collaboratory Living Text book,[27] and other studies that implemented processes to automate the selection of the best NDI match and minimize the number of records requiring manual review.[28–33]

One **distributed program (Program 5)** was designed to import NDI files into SAS software format (refer to the *NDI User's Guide*[18] and technical specifications for Program 5 for specific data file details), apply multiple algorithms for selecting the best NDI match, output files for potential manual review if the best match is unclear (e.g., the algorithm identified multiple NDI records that tie for the selected best match), and compare approaches for selecting the best match to existing Study Site death data. As with previous distributed programs, the Coordinating Center developed Program 5 to distribute to Study Sites. Study Sites run the Distributed Program 5 and return aggregated results to the Coordinating Center. The Coordinating Center prepares a report for the study team to review. **Figure 13** provides an overview of the major functionality provided by Distributed Program 5.

Figure 13. Distributed Program 5: Overview of distributed process for selecting the best match from data returned by the National Death Index (NDI)



Please note, while the NDI provides some pdf summary files, only flat text files are imported. Files used in the process include 'match', 'no match', 'combined', and 'cause' files which are used for matching and quality control purposes; NDI 'combined' and 'cause' files are used provide death information for patients matched to NDI records, please refer to* ***Supplemental Appendix* *for more information.*

A. Importing NDI data

NDI results are returned in "flat text data files," which is a universal format for transmitting data between organizations. Please refer to the *NDI User's Guide* for more details regarding the NDI file structure. [18] These patient-level data are imported into SAS datasets by running Distributed Program 5, and specifications for this program are included in **Appendix B**.

B. Removing duplicate NDI records and excluding poor matches

Remove duplicate death certificates: In our distributed process for selecting the best NDI match, we found it was helpful to identify patients records which were matched to duplicate NDI death records (e.g., the NDI provided the same death certificate number twice for one patient). While our understanding is duplicate records are not common in NDI data, it is important to plan for this potential. Removing any duplicate death certificates provided by the NDI to the Study Sites delivers process efficiencies by removing some potential matches that would need to be evaluated. For example, in situations where the best match is unclear (e.g., there are ties between records in probabilistic matching score or other factors), Program 5 outputs a file for manual review and investigation by Study Sites. By removing any duplicate death certificates provided by the NDI in the data files, we also remove the need to investigate these records manually.

Excluding poor matches: As outlined in **Figure 13**, the next step in the distributed process for selecting the best NDI is to exclude very poor-quality matches. Thus, our process excludes records from being selected as the best match that meet any of the following:

- If the death date in the NDI death record is before the cohort entry date (potential false positive match, where NDI provides a potential matching death record at a time the patient was known to be alive)
- If the probabilistic score ≤0 (negative score indicates many variables did not match)
- If the class code=5 (SSN was available and provided to the NDI but did not match an NDI record)

C. Selecting the best match

Distributed Program 5 creates aggregated tables at each Study Site which are designed to assist with selecting the best match. Study Site tables can be aggregated by the Coordinating Center across Sites.

Aggregated tables describing returned NDI data: Information about the number of patients with at least one match returned by the NDI, the number of NDI matches per patient, and the minimum and maximum number of matches returned per patient is outlined in Table 5 below. This information is intended to help the Study Team understand the volume of data provided by the NDI to the Study Sites, and the number of potential matches that must be evaluated.

Table 5. Number of patients and potential matches returned by the National Death Index (NDI) stratified by criteria used to select individuals to submit to the NDI

Number of patients or potential matches	Death only N (%)*	Death and Disenrolled N (%)*	Disenrolled only N (%)*	LTF N (%)*	Total records submitted to the NDI N (%)*
Patients submitted to the NDI for matching					
Of patients with at least 1 possible match					
Patients with at least 1 match returned N					
Number of matches, per patient (Average, min, max)					
Abbreviations: LTF, Lost to Follow-up **% of patients submitted to the NDI for matching*					

Table 6 is designed to provide more detailed information about variables in the files the NDI returns to the Study Sites. Results are stratified by the NDI variable 'status code', and NDI submission criteria defined by the study team (potential death in Study Site, potential death and disenrollment from health plan, disenrollment from health plan only, or lost to follow up [LTF] in Study Site data). These results assist the Study Team with evaluation of the quality of matches returned by the NDI. Please see Section A above for the definition of each of the NDI provided variables listed in Table 6.

Table 6. Information provided by the National Death Index (NDI) to assist with selecting the best match and algorithm results, stratified by status code and criteria used to select individuals to submit to the NDI

	NDI provided variable, Status code*				Criteria used to select individuals to submit to the NDI for matching									
	Status code= 1		Status code= 0		Death		Death AND Disenrolled		Disenrolled only		Lost to Follow-up		All **	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
NDI Variables														
Exact														
Class 1														
Class 2														
Probabilistic Scores***														
<34.5														
≥34.5														
≥39.5														
≥44.5														
≥49.5														
≥54.5														
Class 3														
Probabilistic Scores***														
<27.5														
≥27.5														
≥32.5														
≥37.5														
≥42.5														
≥47.5														
Class 4														
Probabilistic Scores***														
<22.5														
≥22.5														
≥27.5														
≥32.5														
≥37.5														
≥42.5														
Class 5														

Abbreviations: LTF, Lost to Follow-up;

**Status code is the death status that the NDI suggests. Records with a status code value of 1 are considered true matches and assumed dead; records with a status code value of 0 are considered false matches and are assumed alive. However, the NDI also states that the status code should only be used as a guide. Some potential NDI matches with a status code of 0 may be true matches, some potential matches with status code of 1 may not be true matches.*[18]

***Total submitted to the NDI*

****Please refer to the NDI User's Guide*[18] *for more information about recommended probabilistic match scores stratified by Class*

Aggregated tables applying algorithms for selecting the best match: Distributed Program 5 applies pre-specified algorithms for selecting the best match to returned NDI data. Two automated algorithms utilize NDI's probabilistic matching score and are described as potential approaches for selecting the best match in the *NDI User's guide*.[18] A third automated algorithm utilizes a deterministic matching approach which has successfully been applied by studies in situations where extensive manual review and adjudication of multiple potential matches is not possible.[28–31] Distributed Program 5 applies these algorithms and provides aggregated results for the Study Team to evaluate. **Table 7** is designed to display the number of matches and ties which occur after the application of algorithms 1-3 described below. Results are stratified by the NDI variable 'status code', and NDI submission criteria (Status of death, death and disenrollment, disenrollment only, or lost to follow up [LTF] in Study Site data) and are intended to assist the Study Team with understanding algorithm performance by NDI submission criteria. Please see **Section I.** above for the definition of each of the NDI provided variables listed in **Table 7**.

Table 7. Number of identified single best matches and ties after the application of algorithms 1-3, stratified by status code and criteria used to select individuals to submit to the National Death Index (NDI)

	NDI provided variable, Status code*						Criteria used to select individuals to submit to the NDI for matching														
	Status code= 1			Status code= 0			Death			Death and Disenrolled			Disenrolled only			LTF			All*		
	N	%**	# ties***	N	%**	# ties***	N	%**	# ties***	N	%**	# ties***	N	%**	# ties***	N	%**	# ties***	N	%**	# ties***
Algorithms																					
Algorithm 1																					
Algorithm 2																					
Algorithm 3																					

*Abbreviations: LTF, Lost to Follow-up; * Total submitted to NDI*

**Status code is the death status that the NDI suggests. Records with a status code value of 1 are considered true matches and assumed dead; records with a status code value of 0 are considered false matches and are assumed alive. However, the NDI also states that the status code should only be used as a guide. Some potential NDI matches with a status code of 0 may be true matches, some potential matches with status code of 1 may not be true matches.*[18]

***% of patients submitted to the NDI for matching meeting each algorithm, stratified by Status Code and criteria used to select individuals to submit to the NDI for matching.*

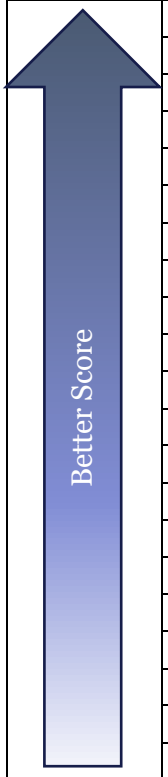
****When two records have identical probabilistic matching scores and match on other variables, these records are considered 'ties'. The number of patients with ties are displayed in this table and are also output into a separate file for manual review and investigation by the Study Sites.*

Description of automated algorithms: Automated algorithms outlined below are pre-programmed options available to future studies. However, as the returned NDI data files are imported into SAS software format and saved at each Study Site, future studies would have the ability to implement a different approach. For example, if a future study wishes to utilize a different approach for selecting the best NDI match, it would be possible to implement a new algorithm within the distributed program.

During the pilot study, we learned that automated algorithm 1 and 2 as outlined below produced identical results. We hypothesize the NDI may use the probabilistic matching score cutoffs recommended in the *NDI User's Guide* to assign status code, but we were unable to confirm. Please keep in mind that future studies could consider modifying the probabilistic matching score cutoffs within algorithm 2 if they wish to implement a more sensitive approach (i.e., evaluate records with a status_code='0').

Automated algorithm 1	Automated algorithm 2	Automated algorithm 3
Utilizing probabilistic matching approach outlined in the NDI User's Guide[1]	*Utilizing probabilistic matching approach outlined in the NDI User's Guide*[1]	*Utilizing deterministic matching criteria to sort potential matches into a hierarchy as outlined by previous studies*[28–31]
• Utilize NDI's variable status code to identify matches • Status_code=1 • Select record with the smallest value of class for the patient • Then within class, select record with the highest probabilistic score among records that meet the NDI's minimum probabilistic score per class[1] (NDI recommended cutoff scores Class 2 score cutoff: ≥44.5, Class 3 score cutoff: ≥37.5, Class 4 score cutoff: ≥32.5), • Save records that tie to a separate file for potential manual review	• Utilize NDI's variable status code to identify matches • Status_code=1 OR 0 • Select record with the smallest value of class for the patient • Then within class, select record with the highest probabilistic score among records that meet the NDI's minimum probabilistic score per class[1] (NDI recommended cutoff scores Class 2 score cutoff: ≥44.5, Class 3 score cutoff: ≥37.5, Class 4 score cutoff: ≥32.5) • Save records that tie to a separate file for potential manual review	• Utilizing deterministic matching criteria to sort potential matches into a hierarchy, based on which specific variables matched. • **Table 8** provides an example of a hierarchical approach that may be adapted to select the single most likely NDI match.[28–31]

Table 8. Midkiff Criteria for sorting National Death Index (NDI) potential matches into a hierarchy to facilitate selection of a single most likely NDI match[28,29,31]

Score	Criteria
	SSN, name, sex, elements of DOB
	SSN, elements of name (NYSIIS), sex, DOB
	SSN close (>6 digits match), elements of name (NYSIIS), sex, and DOB
	SSN close, elements of name (NYSIIS), sex, elements of DOB
	SSN, first name, last name, sex
	SSN, elements of name (NYSIIS), sex
	SSN, last name, sex, elements of DOB
	SSN, first name, last name, DOB
	SSN, first name, last name, elements of DOB
	SSN close, first name, last name, birth M, D, Y
	SSN, first name, OR last name, sex, birth M, D, Y
	SSN, sex, birth M, D, Y, and demographics *
	SSN unknown, name (not common), sex, DOB, demographics*
	SSN close, name (not common), OR elements of name (NYSIIS), sex, DOB
	SSN unknown, name (not common), sex, DOB
	SSN unknown, name (very rare), sex, birth M, D, Y+_ 3
	SSN, first name OR last name OR 2 of 3 elements of DOB, sex
	SSN, first name
	NDI status = true match (assumed dead) and class = 2, 3, 4
	SSN unknown, name (middle initial not missing), sex, DOB
	SSN, sex

D = day; M = month; NYSIIS= New York State Identification and Intelligence System phonetic code; Y= year, SSN =Social Security Number, DOB = Date of Birth
**Demographics = race, marital status, state of birth*

D. Outputting files for manual review when best match in unclear

In situations where the best match is unclear (e.g., there are ties between records in probabilistic matching score or other factors) within each of the three algorithms, a file is output to provide the potential for manual review and investigation by Study Sites. After algorithms 1-3 have been applied, we generally expect the volume of records in which the best match is unclear or where there are ties to be small, but volume of ties could vary by study. In our pilot study, we defined a tie between two potential NDI matches as two records with an exact probabilistic matching score and ultimately did not identify any 'ties'. However, subsequently we learned that some studies also consider relaxing the definition of 'ties' to include potential matches within a range of the probabilistic matching scores (rather than requiring an exact match). Using this less stringent approach may identify more 'tied' NDI records, which would require additional manual evaluation by the Study Sites, but also provides opportunities for confirming match selection in situations where multiple potential matches have very similar probabilistic matching scores. In future studies, Program 5 could be adjusted to include a broader definition of 'tied' NDI records, as needed.

If manual review and investigation by Study Sites of records is not possible, the following automated rules could be considered to resolve situations in which the best match is unclear:

- Exclude all ties, i.e., exclude patient from analysis (death status =unknown).
- Assume dead (death=yes), but death date=unknown, truncate follow up at earliest death date, and do not use cause of death information.

We anticipate different Study Teams may choose to implement different approaches when the best match is unclear, and that reviewing study tables output by distributed program 5 may provide useful information to assist with determining an approach for resolving ties. Thus, distributed program 6 provides an opportunity to customize rules for resolving situations where the best match in unclear**.**

E. Examining concordance between the NDI results and study site death information

To facilitate decision making, it may be helpful for Study Teams to evaluate death data received from the NDI. Five of the six Study Sites included in the pilot project had already matched their databases to state death records. This provided an opportunity to build an indirect validation step within distributed program 5 which utilizes death information available to the Study Sites and examine concordance between NDI results and Study Site data. This option is available to future studies and requires the Study Team to specify the time periods in which death information is considered well populated within each Study Site (e.g., Years 2000 through 2015).

When this time-period is specified, **Distributed Program 5** returns the concordance of Study Site death information with information attained through linkage to NDI data. The program assumes Study Site death data are the gold standard and results attained by applying algorithms 1-3 to returned NDI data are the comparator. Metrics such as sensitivity (SEN), specificity (SPEC), positive predictive value (PPV), and negative predictive value (NPV) are output for the Study Team to evaluate. **Table 9** is designed to describe concordance between deaths identified at each Study Site and through the NDIstratified by the NDI provided status code and the submission criteria used to select individuals to submit to the NDI for matching.

Future studies could also consider utilizing medical records, autopsy reports, death certificates, ambulance, or other similar records to validate NDI death information, if available.

Table 9. Concordance between Study Site and National Death Index (NDI) deaths identified, stratified by status code and criteria used to select individuals to submit to the NDI

	True Positive (a)	False Positive (b)	False Negative (c)	True Negative (d)	Test Positive (a+b)	Test Negative (c+d)	Death in Study Site (a+c)	No Death in Study Site (b+d)	Total population (a+b+c+d)	SEN	SPEC	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Algorithm 1 Total*													
NDI Variables													
Exact													
Class 1													
Class 2													
Scores													
<34.5													
≥34.5													
≥39.5													
≥44.5													
≥49.5													
≥54.5													
Class 3													
Scores													
<27.5													
≥27.5													
≥32.5													
≥37.5													
≥42.5													
≥47.5													
Class 4													
Scores													
<22.5													
≥22.5													
≥27.5													
≥32.5													
≥37.5													
≥42.5													
Class 5													

Abbreviations: SEN, Sensitivity; SPEC, Specificity; PPV, Positive Predictive Value; NPV, Negative Predictive Value

**Please note, this table will be repeated for algorithms 2 and 3.*

III. Saving the best match for incorporation into study analytic datasets

After reviewing the aggregate data returned by Distributed Program 5 and determining the approach for selecting the 'best' NDI match from all NDI matches, the Study Team will need to implement a program that will save the final match for study analyses. The chosen rule(s) may be to simply apply one of the three algorithms or may include additional requirements such as requiring exact matches on selected variables. The application of these Study Team developed rules occurs within **Distributed Program 6**. The specifications we used in the pilot project for saving the final matches for study analyses are included in a **Supplemental Appendix**, but please keep in mind this is a prototype that can be reused and modified for application to future studies.

In addition to applying the rules, **Distributed Program 6** will also structure the NDI data into data files that will be most useful for analyses. These files will include selected data from Program 1 (cohort identification) and future studies may choose to combine these data with other variables required for final study analyses include other (e.g., exposures, covariates, enrollment time, etc.) that will be used in final study analyses. We provide more information about the structure used to save the final NDI matches for use in the pilot project in **Chapter 5.**

Chapter 5: Creating Study Tables for Linked NDI Death Information

After the study team chooses an algorithm to select the "best" NDI match as described in **Chapter 4**, the next step of our distributed linkage process applies the selected NDI algorithm and saves NDI records in a uniform format for final analyses. A uniform format streamlines analyses in a multi-site research study and enables a single distributed SAS program to be executed across sites without requiring site-specific program modifications.

Distributed Program 6 implements a uniform format for retaining death information attained from the NDI and creates two tables, one containing NDI death information and one containing NDI cause of death information. This structure leverages and expands upon the Sentinel Common Data Model's Death and Cause of Death tables,[34] which were utilized by the pilot study described in previous chapters.

In order to provide flexibility to future studies, we prioritized ensuing all information attained from the NDI is saved in a usable format, regardless of whether the data element was used during the pilot study. For example, although the pilot study did not include infants, we prioritized flexibility and reusability and built the table structure to accommodate all NDI information about causes of death that were reclassified within infants (See **Table B5** and the **NDI Users Guide**[18] variable 'Infant Cause Recode', Exhibit 8 for more details). However, we would recommend additional investigation exploring the efficiency and utility of including all specific NDI variables if or when these tables are considered for incorporation into other common data models (e.g., Sentinel Common Data Model, PCORnet Common Data Model) for more routine use. In addition, before incorporating this table structure into other common data models we would also recommend exploring the use of either a project-specific or IRB-specific study number saved locally at each study site to facilitate data management and tracking of simultaneous multi-site studies. The tables are named by Distributed Program 6 as follows:

NDI Death Table (NDI_Death): This table captures *fact of death* information for a decedent.

NDI Cause of Death Table (NDI_COD): This table captures *cause of death* information for a decedent.

Information about the data returned by the NDI can be found in the *NDI User's Guide*, Chapter 3.[18] The structure of death and cause of death tables are included in **Appendix B**. These tables separate identifiable NDI information from other study analytic files, which facilitates the data disposition process as outlined by the NDI in the data disposition section of the NDI application (i.e., in most cases all identifying or identifiable data received from NDI must be removed from all research records at the conclusion of the study or within 5 years after receipt of the NDI data – regardless of the data set in which the data are kept). We recommend that future studies keep this NDI policy in mind when creating study tables for linked NDI death information and consult the data disposition section of the NDI application and with the NDI to ensure policies are followed.

References

1. Toh S, Platt R, Steiner JF, Brown JS. Comparative-Effectiveness Research in Distributed Health Data Networks. *Clin Pharmacol Ther*. 2011;90(6):883-887. doi:10.1038/clpt.2011.236

2. Maro JC, Platt R, Holmes JH, et al. Design of a national distributed health data network. *Ann Intern Med*. 2009;151(5):341-344. doi:10.7326/0003-4819-151-5-200909010-00139

3. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. *Med Care*. 2010;48(6 Suppl):S45-51. doi:10.1097/MLR.0b013e3181d9919f

4. Fuller C, Hau W, Leonard C, et al. Developing a standardized and reusable method to link distributed health plan databases to the National Death Index: a methods development study protocol. *JMIR Res Protoc*. doi:10.2196/21811

5. About NCHS - NCHS Fact Sheets - National Death Index. Published June 1, 2019. Accessed October 22, 2020. https://www.cdc.gov/nchs/data/factsheets/factsheet_ndi.htm

6. Federal Policy for the Protection of Human Subjects ('Common Rule. HHS.gov. Published June 23, 2009. Accessed October 22, 2020. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

7. Rights (OCR) O for C. Disclosures for Public Health Activities. HHS.gov. Published January 7, 2009. Accessed December 15, 2020. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/disclosures-public-health-activities/index.html

8. hhs.gov. Human Subject Regulations Decision Charts. HHS.gov. Published February 25, 2016. Accessed October 22, 2020. https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts/index.html

9. Rights (OCR) O for C. Research. HHS.gov. Published January 7, 2009. Accessed October 22, 2020. https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/research/index.html

10. Rosati K, Jorgensen N, Plc CB. HIPAA and Common Rule Compliance in the Sentinel Initiative. :49.

11. Pyle S. Benefits of Working with a Central IRB, Improved Efficiencies and Enhanced Human Subject Protections. Published online 2013.

12. SMART IRB | National IRB Reliance Initiative. Accessed October 22, 2020. https://smartirb.org/

13. SMART IRB: Master Common Reciprocal Institutional Review Board Authorization Agreement Standard Operating Procedures.pdf. Accessed October 22, 2020. https://smartirb.org/assets/files/SMART_IRB_SOP-090816.pdf

14. How to Use the National Death Index: Steps in the Process. Accessed December 22, 2020. https://www.cdc.gov/nchs/ndi/portal.htm

15. NOT-OD-20-057: Notice of Information: National Death Index Linkage Access for NIH-Supported Investigators. Accessed December 4, 2020. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-057.html

16. *SAS and All Other SAS Institute Inc. Product or Service Names Are Registered Trademarks or Trademarks of SAS Institute Inc. in the USA and Other Countries. ® Indicates USA Registration.*

17. Routine Querying System. Sentinel Initiative. Published August 16, 2016. Accessed October 22, 2020. https://www.sentinelinitiative.org/methods-data-tools/routine-querying-tools/routine-querying-system

18. National Dealth Index User's Guide. Published online 2013. https://www.cdc.gov/nchs/data/ndi/ndi_users_guide.pdf

19. Social Security History. Accessed December 14, 2020. https://www.ssa.gov/history/ssn/misused.html

20. SSA - POMS: RM 10201.035 - Invalid Social Security Numbers (SSNs) - 06/23/2011. Accessed December 14, 2020. https://secure.ssa.gov/poms.nsf/lnx/0110201035

21. WinSCP :: Official Site :: Free SFTP and FTP client for Windows. Accessed December 14, 2020. https://winscp.net/eng/index.php

22. Data Access - National Death Index - Early Release Program. Published February 28, 2019. Accessed October 22, 2020. https://www.cdc.gov/nchs/ndi/ndi_early_release.htm

23. Fellegi IP, Sunter AB. A Theory for Record Linkage. *J Am Stat Assoc*. 1969;64(328):1183-1210. doi:10.1080/01621459.1969.10501049

24. Ball R, Robb M, Anderson SA, Dal Pan G. The FDA's sentinel initiative--A comprehensive approach to medical product surveillance. *Clin Pharmacol Ther*. 2016;99(3):265-268. doi:10.1002/cpt.320

25. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System — A National Resource for Evidence Development. *N Engl J Med*. 2011;364(6):498-499. doi:10.1056/NEJMp1014427

26. Weeks J, Pardee R. Learning to Share Health Care Data: A Brief Timeline of Influential Common Data Models and Distributed Health Data Networks in U.S. Health Care Research. *eGEMs*. 7(1). doi:10.5334/egems.279

27. Eisenstein E, Curtis L, Prather K, Hernandez A, Weinfurt K. NIH Collaboratory Living Textbook of Pragmatic Clinical Trials: Choosing and Specifying Endpoints and Outcomes. Rethinking Clinical Trials. Accessed October 23, 2020. https://rethinkingclinicaltrials.org/chapters/design/choosing-specifying-end-points-outcomes/choosing-and-specifying-endpoints-and-outcomes-introduction/

28. Identifying causes of death in a large cohort of asthma patients. RTI-HS. Published November 26, 2019. Accessed October 22, 2020. https://www.rtihs.org/publications/identifying-causes-death-large-cohort-asthma-patients

29. Midkiff KD, Calingaert B, Tennis P, Harris D, Andrews EB. Using an Automated Computer-Based Algorithm to Increase the Efficiency of Selecting the Best Match From National Death Index Linkage Results. :1.

30. Johannes CB, McQuay LJ, Midkiff KD, et al. The feasibility of using multiple databases to study rare outcomes: the potential effect of long-acting beta agonists with inhaled corticosteroid therapy on asthma mortality. *Pharmacoepidemiol Drug Saf*. 2017;26(4):446-458. doi:10.1002/pds.4151

31. Wojcik NC, Huebner WW, Jorgensen G. Strategies for using the National Death Index and the Social Security Administration for death ascertainment in large occupational cohort mortality studies. *Am J Epidemiol*. 2010;172(4):469-477. doi:10.1093/aje/kwq130

32. Binswanger IA, Morenoff JD, Chilcote CA, Harding DJ. Ascertainment of Vital Status Among People With Criminal Justice Involvement Using Department of Corrections Records, the US National Death Index, and Social Security Master Death Files. *Am J Epidemiol*. 2017;185(10):982-985. doi:10.1093/aje/kww221

33. Cleveland C, Crawford P, Fuoco M, Ng D, Yan D. Death Data in and Beyond in the Health Care Systems Research Network. *Health Care Syst Res Netw 2019*. Published online 2019.

34. Sentinel Common Data Model. Sentinel Initiative. Published February 2, 2016. Accessed October 22, 2020. https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model/sentinel-common-data-model

35. Network THCSR. The Health Care Systems Research Network Website. Published 2015. http://www.hcsrn.org/en/

36. Our Mission. Accessed October 23, 2020. https://www.aahrpp.org/learn/about-aahrpp/our-mission

37. Chung CP, Murray KT, Stein CM, Hall K, Ray WA. A computer case definition for sudden cardiac death. *Pharmacoepidemiol Drug Saf*. 2010;19(6):563-572. doi:10.1002/pds.1888

38. Zheng ZJ, Croft JB, Giles WH, Mensah GA. Sudden cardiac death in the United States, 1989 to 1998. *Circulation*. 2001;104(18):2158-2163. doi:10.1161/hc4301.098254

39. Carnes C. Antiarrhythmic drug classifciaton. In: *Novel Therapeutic Targets for Antiarrhythmic Drugs*. Vol In: Billman GE, ed. 1st Edition. Wiley; 2010.

40. Vaughan Williams E, Cambell T ed. Classification of antiarrhythmic actions. In: *Antiarrhythmic Drugs*. 1st Edition. Springer; 1989.

41. NVSS - MMDS - ACME and TRANSAX. Published March 4, 2019. Accessed October 22, 2020. https://www.cdc.gov/nchs/nvss/mmds/acme_transax.htm

42. Underlying and Multiple Cause of Death Codes. NCHS Surveys, 2011 Linked Mortality Files. Accessed October 22, 2020. https://www.cdc.gov/nchs/data/datalinkage/underlying_and_multiple_causes_of_death557_2011.pdf

43. National Center for Health Statistics. Mortality Medical Data System (MMDS) Software Documentation. Accessed October 22, 2020. https://www.cdc.gov/nchs/data/dvs/helpfile.pdf

44. National Center for Health Statistics. U.S. Standard Certificate of Death. Accessed October 22, 2020. https://www.cdc.gov/nchs/data/dvs/DEATH11-03final-acc.pdf

Appendix A: Supplemental Materials for Administrative Workflow

I. Process for Achieving Central IRB Approval

This document is intended to supplement the Central IRB process information in **Chapter 1**, provide additional considerations for selecting a Central IRB, and outline roles and responsibilities for each collaborating team (i.e., Project Lead Team, Coordinating Center, Study Sites, and Central IRB).

A. Considerations for selecting a Central IRB site

The Central IRB of record has regulatory responsibility for assuring the protection of the rights and welfare of welfare of research participants from initial review to termination of the research, including review and approval of a waiver of informed consent. This responsibility will drive main considerations for the selection the lead IRB for the Central IRB process, which include:

1. Willingness and ability to serve as a Privacy Board to fulfill the requirements of the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule for use or disclosure of protected health information for research;
2. Adherence to communication standards and a commitment to transparency through sharing information about the review process;
3. Appropriate expertise and experience to review the proposed research and the capacity to review the study protocol and participating sites;
4. Familiarity with Central IRBs process in general and specifically those available within multi-site Clinical Trial Networks may be helpful. Examples used in the pilot project, include Health Care Systems Research Network[35] and SMART IRB,[12] a network established as part of NIH single IRB Review policy (effective date: January 25, 2018).
    a. These networks include a streamlined Central IRB process, including a Master Reliance Agreement, and an online reliance system that allows institutions to request, track, and document reliance arrangements and eliminates the need to sign separate agreements on a study-by-study basis.
5. Accreditation from the Association for the Accreditation of Human Research Protection Program (AAHRP).[36] Indicates the IRB organization follows rigorous standards for ethics, quality, and protections for human research. Although not required, sites may prefer this accreditation.

B1. Overview of Lead IRB's Central IRB Responsibilities

Responsibilities (non-protocol-specific):

1. Maintain program for education and training in human subjects research for IRB personnel.
2. Must be registered with FDA and OHRP.
3. Notify study site institutions if accreditation status changes.

Responsibilities (protocol-specific):

1. Ensure the study meets generally accepted ethical standards of human subject's protections and complies with applicable regulations, for example, the Common Rule (45CFR 46)[9] as well as state and applicable international regulations, such as the European Clinical Trial Directive.
2. Collect, review, and consider site-specific information provided by the individual sites. This information could include special considerations regarding state laws and any restrictions placed on the study by the institution, such as feasibility of the research or special training requirements.

3. Review and approve the HIPAA waiver of informed consent form (if applicable) and any other research-related documents.
4. Provide the Study Site with copies of IRB approval documents, IRB rosters, and meeting minutes upon request or in accord with the IRB authorization agreement.
5. Notify the Study Site promptly in writing of serious or continuing non-compliance or unanticipated problems involving risks to subjects or others.
6. Notify the Study Site promptly in writing of any suspension or termination of central IRB approval and of remedial actions required of the institution or its agents by the central IRB. If review is for an institution that conducts federally-funded research, the central IRB must commit to adhere to the requirements of the institution's federal-wide assurances (FWA(s)).

B2. Overview of Study Site responsibilities

Responsibilities (non-protocol-specific):

1. Maintain education of investigators and research staff and training in human subjects research.
2. Maintain policies and procedures for the conduct of human subjects research as appropriate for the institution.
3. Maintain appropriate institution-specific required credentialing of staff.
4. Maintain approved federal-wide assurances (FWAs), including ensuring that the arrangement with the central IRB is documented by a written agreement.
5. Conduct a privacy and security review as required by the Health Insurance Portability and Accountability Act (HIPAA) with respect to the mechanisms for permitting the use and disclosure of Protected Health Information (PHI) for research
6. Ensure that the site Principal Investigator and staff are conducting research in accordance with IRB-approved protocol, procedures, and documents.

Responsibilities (protocol-specific):

I. Designate the IRB of record for the protocol.
II. Obtain IRB approval of research protocols involving human subjects.
III. Notify the IRB promptly in writing of serious or continuing non-compliance or unanticipated problems involving risks to subjects or others.
IV. Evaluate the local context in which the research will be conducted, including consideration of any specific requirements of state or local laws, regulations, policies, or standards. Participating Study Sites should inform the central IRB of any relevant requirements or findings from the analysis that would affect conduct of the research at that institution.
V. May provide a waiver of authorization as described under the Health Insurance Portability and Accountability Act (HIPAA), if applicable to the specific study (i.e., not all studies will request a waiver).

For more information regarding Standard Operating Procedures for using a Central IRB, visit Streamlined, Multisite, Accelerated Resources for Trials (SMART) IRB: Master Common Reciprocal Institutional Review Board Authorization Agreement Standard Operating Procedures: (https://smartirb.org/assets/files/SMART_IRB_SOP-090816.pdf). The SMART IRB platform is a national resource for IRB review of multisite studies.

II. Completing a National Death Index (NDI) Application

The completion of the NDI application is a multi-step process that requires Coordinating Center, Sponsor and Study Site legal, privacy, and IT security reviews and timing coordination. Below are the steps and correspondence document examples.

For efficiency we recommend the Coordinating Center initiate the draft application. We provide detailed instructions for completing each section of the application, based on lessons learned from the pilot process and information gathered from the NDI. If a description does not fit in the question box provided on the NDI provided PDF form, the NDI requests a separate document to be included as an attachment to the application.

1. **Drafting the main NDI application:** Recommendation for each section of the NDI application are included in this section and will need to be completed as follows:
   - **Sections 1-5:** Cover page of the National Death Index Application Form with title study, individual and organization requesting use of the NDI, Co-Principal Investigators, and Type of NDI search requested. The type of NDI Search requested will require the result from a feasibility analysis and should be requested as UNKNOWN study status of study subjects. The external funding source should also be included.
   - **Section 6. Data Sources:** To complete the data source section of the application, list each organization that will be requesting the data from NDI and explain the ***type of data*** will be collected for the study. Include a succinct overview of the database structure and the acquired longitudinal data collected at Study Sites. In the pilot study we chose to include language that no identifiers, including dates, will be returned to the Coordinating Center and provided a brief description of the NDI request for death and cause of death data. Other studies might choose a different approach.
   - **Section 7. External Organization (other than the NDI applicant's organization) receiving Identifying or Identifiable information:** This section must list each Study Site separately and clearly define their role and list the activities that will be performed at Study Sites. It is important to note where any PHI information will be maintained. In the pilot study, we stated that PHI information would be maintained behind each of the Study Site's firewalls, and would not be shared with other Study Sites or the Coordinating Center. Do not include the Coordinating Center in this section unless the Coordinating Center is also a Study Site requesting to link to the NDI.
   - **Section 8: Summary of Study Protocol or Project Activities:** The NDI recommends including a high-level overview of the planned study including the population, study aims, study design and background, sources of data, study population, eligibility criteria, study duration and how long the NDI data will be used. ***This section should succinctly describe the purpose of linking to the NDI.*** A supporting protocol is not required to be submitted to the NDI and language included in this section should be void of complex scientific methods. The provided context should be understood by an audience without specific knowledge of the research that will be conducted.
   - **Sections 9-11*:*** During the pilot project, we selected no follow-back investigations or death certificates option (i.e., we agreed that we would not conduct follow-back investigations or retrieve death certificates). For the IRB section, list the date of the initial IRB application approval for the Coordinating Center. It is important to note that the NDI will require a letter from each Study Site's IRB in addition to the Coordinating Center IRB approval letter at the time of the final

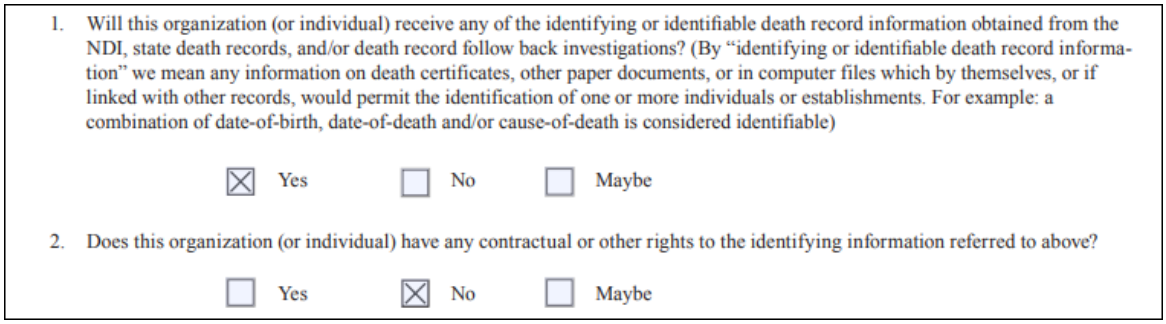
NDI application submission. The IRB letters must contain the same project title and CO-PIs listed in the cover page.

- **Section 12:** This section should delineate the sites submitting records to the NDI and receiving results from the NDI search. Section 12b. should list the physical, technical, and administrative controls for the Coordinating Center and Analytic Center. Study Site physical, technical, and administrative controls will be covered in the Supplemental Confidentiality Agreements.
- **Section 13: Data Disposition:** It is critical to ensure that the correct data disposition date is included. While the NDI notes in the application that destruction of the PHI data may occur at the completion of the study or within five years after receipt, the NDI reviews the date include in box #1 as the date of the final application submission. If the date is more than 5-years after the application submission date, then it is critical to provide assurance that an extension request will be made, or the certificate of data disposal will be submitted to the NDI no later than 5-years following the date of the final application submission.
- **Sections 14-15: Study Completion and Other Uses of the Data:** The study termination should be provided, and no other uses of the date should be selected.
- **Section 16: Types of Data to Be Submitted to the NDI:** This section will require collaboration with participating Study Sites to estimate the percentage of records that contain the PHI criteria used to link to the NDI. These percentages will help NDI maximize the number of true matches identified and will assist the Study Sites and Coordinating Center assess the quality of the matches.

2. **Executing Confidentiality Agreements for inclusion in the NDI application:** The NDI will require the Coordinating Center to sign a confidentiality agreement for the study confirming agreement to the terms and conditions associated with the NDI application and to the use of the information obtained from the NDI. It is the Coordinating Center's responsibility to ensure terms and conditions are followed throughout the study for all Study Sites included in the application. There are three signatures included on this agreement form (1) Data Steward (typically the Study Site Data Manager)(2) Principal Investigator and (3) Official authorized to execute agreements. If the Coordinating/Analytic Center is not requesting data from the NDI, then the Data Steward section should not be completed and signed. Signatures for this form must be signed by the Principal Investigator followed by the authorized official ***in wet ink*** (electronic signatures will not be accepted). In addition, signatures MUST be signed and dated in following order:
    - **i.** Data Steward
    - **ii.** Principal Investigator
    - **iii.** Official authorized to execute agreements (last to sign)
3. We recommend the Coordinating Center's grant or legal department has an opportunity review the draft application and Confidentiality Agreement before obtaining signatures. Once approval is received from grant/legal department, signatures may be collected while the draft is under review.
4. The study Sponsor will also be required to submit a Confidentiality Agreement and should be collected concurrently with the Coordinating Center. The same process should be followed; however, the NDI accepted an electronic signature for the FDA during the pilot project. Other Sponsors should work with the Coordinating Center to confirm if electronic signatures will be accepted by the NDI.

B. Instructions for review of draft NDI application by Study Sites and Sponsor.

1. **Review Draft Application:** Once the NDI application is drafted by the Coordinating Center, the application should be shared with Study Sites and Sponsor for their initial feedback and opportunity to address any concerns.
2. **Draft Supplemental Confidentiality Agreements:** In parallel to the review of the main application, Study Sites will complete a draft Supplemental Confidentiality Agreement. The NDI thoroughly reviews these draft agreements before collecting any signatures to ensure appropriate physical, technical, and administrative controls are in place to protect the NDI data exchange with Study Sites. ***Language must be included that NDI files received by Study Sites will be kept in a dedicated, NDI secure folder, separate from administrative datasets/records, will only be accessed by a study team member in the IRB application and will not be combined with any non-study administrative datasets/records***. The Coordinating Center should provide draft recommended language to Study Sites; however, each Study Site will need to confirm their specific physical, technical, and administrative controls. Sites should also select the following:



1. Will this organization (or individual) receive any of the identifying or identifiable death record information obtained from the NDI, state death records, and/or death record follow back investigations? (By "identifying or identifiable death record information" we mean any information on death certificates, other paper documents, or in computer files which by themselves, or if linked with other records, would permit the identification of one or more individuals or establishments. For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

☒ Yes ☐ No ☐ Maybe

2. Does this organization (or individual) have any contractual or other rights to the identifying information referred to above?

☐ Yes ☒ No ☐ Maybe

3. We recommend the Coordinating Center provide Study Sites with a deadline to draft Supplemental Confidentiality forms (e.g. 2-3 weeks); this will ensure a draft application is submitted to the NDI as soon as possible.

C. Instructions for submitting a draft NDI application to NDI, and soliciting feedback from NDI advisors

1. The draft NDI application including draft **unsigned** Supplemental Confidentiality Forms and Coordinating Center signed Confidentiality form, should be submitted to the NDI via email to ndi@cdc.gov. While signed documents can be submitted to the NDI for review, NDI may require revisions to text in some of these documents and in these situations the agreements would need to be resigned. See **Section III. NDI Initial Application Submission - Template Correspondence** for example language to include in the draft email. It also is recommended to submit the initial IRB approval for the study while waiting on the IRB cede letters from each of the Study Sites.
2. The Coordinating Center may receive an 'acknowledgement' email from the NDI with an NDI Application number (e.g. **NDI 2018-0010**) see **Section IV. NDI Acknowledgement Correspondence**. This application number should be referenced in any communication with the NDI and the cover page of the NDI application should be revised to include the application number. This process may have changed with the introduction of the **NDI portal**, so future studies should also consult with the NDI regarding application tracking methods.

DEPARTMENT OF HEALTH AND HUMAN SERVICES Public Health Service Centers for Disease Control and Prevention National Center for Health Statistics	Assigned NDI Application Number **2018-0010**

3. The Coordinating Center should follow-up with NDI within 5 business days to confirm receipt of the draft NDI application and to confirm when the application will be reviewed for feedback.
4. The Coordinating Center should discuss, and document the NDI required edits to the draft application.
5. The Coordinating Center will work with Study Sites to make any required changes to the Supplemental Confidentiality agreements and send the revised draft to NDI for approval to proceed with signatures.
6. Following the revisions, the application should be shared with the Sponsor for review and any final edits should be incorporated.

D. Instructions for submitting a final NDI application to the NDI board for review

1. *Note: We developed the following guidelines for submitting a final paper NDI application during the pilot study and before the **NDI portal** launch. We recommend checking with the NDI to confirm the process for submitting the final NDI application.* Following NDI acceptance to proceed with Supplemental Confidentiality Agreement signatures, the Coordinating Center will send instructions to the Study Site to execute agreements and include a deadline for sending an electronic copy via email and a hard copy via mail. See **Section V** for example instructions.
2. The final NDI application should be prepared with supporting documents including:
   a. PDF NDI application (unsigned)
   b. Scanned copies of IRB letters. It is possible that all IRB letters might not be available depending on timing. This is allowable to the NDI team, however, important to include in the final submission correspondence to list site IRB letters that are pending.
   c. Scanned signed copies of all Confidentiality Agreements (these should be packaged separately from the main NDI application).
   d. Do not mail the original copy of the NDI application unless instructed by the NDI and confirm in writing that the original copy should be sent if requested.
3. A copy of the final NDI application should be provided to the Study Sites and Sponsor for their records. In some Institutions, a Grants office or Legal review may also be required before the final NDI application can be submitted.
4. See **Section VI** for example instructions to submit the final application to the NDI.
   a. Following final submission, the Coordinating Center should follow-up regarding the status of the application and timeline for NDI review board.

III. NDI Initial Application Submission – Template Correspondence

To: ndi@cdc.gov

Subject Line: NDI Application Form [Unsigned – Request to Open New Project with NDI]

To Whom It May Concern:

Accompanying this correspondence, please find an attachment including The National Death Index Application on behalf of [Institute]. This application is for a new project titled ["Project Title"].

Please accept this submission as an initial draft of a new request for your review and consideration as well as a request to generate an assigned NDI number.

This submission is ahead of an application form with all the necessary signatures for confidentiality agreements, data disposition and IRB approval, with the goal of ensuring we have adequately addressed the project parameters and NDI requirements before completing for final submission.

We look forward to working with you regarding our application and please reach out with any questions.

Best Regards,

IV. NDI Initial Application Submission – NDI Acknowledgement Correspondence

From:
Subject: Acknowledgment of National Death Index (NDI) Application 2018-0010

I am writing to acknowledge receipt of your application 2018-0010 "A reusable, generalizable method to link health plan data with the National Death Index to examine the associations between medical products and death and causes of death".

Your application will be reviewed by an NDI staff member. You will be notified if any changes/clarifications are needed. If you are required to submit clarification/changes to your application, please be sure to cite your application number 2018-0010 when submitting any change.

If your application is approved you will receive notification where to send your data file. You can find the fee schedule and fee worksheet on our website http://www.cdc.gov/nchs/ndi.htm

The NDI staff strongly recommends that all applicants refer to Chapter 2 -Preparing Your Records for instructions and guidelines on setting up the data file in preparation of sending it to the NDI for a match against the database. This 65 page document replaces the following two NDI documents: *NDI User's Manual* and the NDI Plus: Coded Causes of Death.

In all future communications pertaining to this NDI application, please make reference to your assigned NDI Application Number 2018-0010. If you have questions, please contact [Name] at xxx-xxx-xxxx. You may also reach her via email at [email].

Thank you for allowing the National Death Index help you with your research needs.

V. NDI Initial Application Submission – Supplemental Confidentiality Agreement

Executing Signatures for Supplemental Confidentiality Agreements

Due Date:

IMPORTANT TO NOTE:
Application must be signed in the following order in wet ink (we are requesting signatures in blue ink). NDI has strict requirements regarding signatures and will not be accepted if requirements are not followed.

- ***1st Signature*** Data Steward (*who will act as the custodian of the NDI files and will be responsible for observance of conditions of use)*
- ***2nd Signature***: Principal Investigator
- ***3rd Signature:*** Official authorized to execute agreements

Submission: Please send both electronic AND hard copies (see addresses below)

ELECTRONIC	Coordinating Center Point of Contact (POC) email
HARD COPY	ATTN: POC Title Address Phone Fax

VI. NDI Final Application Submission – Electronic Submission

To:

Subject: 2018-0010 Application Update | Final Application & Signed Supplemental Confidentiality Agreements

Hello [Name],

I am attaching Harvard Pilgrim Health Care Institute's ***final*** NDI application (No. **2018-0010**), completed based on your feedback and recommended revisions. I am also attached fully executed NDI Supplemental Confidentiality Agreements from our six participating data partners as a separate PDF package as requested as well as FDA'S electronically signed SCA as permitted.

We will express mail signed NDI application form and all IRB approval documents as instructed, but please let us know if there is anything else that is needed to ensure compliance to proceed with the next steps for panel review.

Thank you,

Attachments

1. *NDI Application & IRB Letters from all Study Sites*
2. *Legal Agreements*

VII. Example NDI Application

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control and Prevention
National Center for Health Statistics

Assigned NDI Application Number

2018-0011

NATIONAL DEATH INDEX APPLICATION FORM

1. **Title of Study or Project** (Must match IRB)

2. **Individual and Organization Requesting Use of NDI**

Principal Investigator or Project Director: Lead Principal Investigator
Title: Title
Organization: Organization

Complete mailing address: (include street address, room number, city, state, and zip code)
Mailing Address
City, State, Zip Code

Phone no.: Ext: E-mail:

Who should be contacted if more information is needed?:

Phone no.: Ext: E-mail:

3. **Co-Principal Investigators** (if any): If there are no Co-PIs, type "None." (Co-PIs employed by the above organization must complete and sign the Confidentiality Agreement. Co-PIs in other organizations must complete and sign the Supplemental Confidentiality Agreement.)

Name(s)	Organization(s)	Phone number(s)
Study Site Principal Investigator #1 Study Site Principal Investigator #2 Study Site Principal Investigator #3	Study Site #1 Study Site #2 Study Site #3	xxx-xxx-xxxx xxx-xxx-xxxx xxx-xxx-xxxx

4. **Type of NDI Search Requested**

		Estimated number of records to be submitted
Routine **NDI file search only**		
NDI *Plus* coded causes of death	Status of study subjects **UNKNOWN**	~145,000
NDI *Plus* coded causes of death	A separate file of **KNOWN** decedents	

1

3.

Name(s)	Organization(s)	Phone number(s)
Study Site Principal Investigator #4	Study Site #4	xxx-xxx-xxxx
Study Site Principal Investigator #5	Study Site #5	xxx-xxx-xxxx
Study Site Principal Investigator #6	Study Site #6	xxx-xxx-xxxx

5. External Funding Sources (If none, type "Internal funding only.")

List the names of all OTHER organizations providing funding for this project and indicate the type of support provided; i.e., grant, contract, cooperative agreement, interagency agreement, other (specify). (NOTE: Except for a FEDERAL GRANT, each sponsor must complete and sign an NDI Supplemental Confidentiality Agreement at the end of this application form.)

Names of Organization(s)	Type of Funding Support
List Funder	Contract (#ABCDEF1234)

6. Data Sources

List all organizations (including your own) which have collected (or will be collecting) data on the study subjects. Under each organization listed, describe the types of data collected. If any of the **external** organizations listed will be receiving identifying or identifiable death record information, they must also be listed in item 7 below. **"Identifying or identifiable death record information"** refers to any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identification numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual.

See insert on next page.

NOTE: Attach additional page if necessary. -

6. Data Sources

Organizations involved in this application that will be collecting data are:

Study Site #1, Study Site #2, Study Site #3, Study Site #4, Study Site #5, and, Study Site #6

All organizations listed above will be using data already available to them through the normal course of business, including administrative claims data as well as data from electronic health records. These databases are formatted into a standard structure (Common Data Model) which uses standard variable names and coded values. Because all the databases are formatted in the same manner, the study coordinating center at [Lead Study Site] can write one program which is then run by each data partner against its local database, returning de-identified data to the coordinating center via a secure system. The organizations comply with standards established by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the Federal Information Security Management Act of 2002 (FISMA).

This study is an IRB approved research activity. Six data partners (aka health plans) have joined this NDI + linkage study and have acquired longitudinal data from inpatient and outpatient settings within their local health plan systems including patient demographics, diagnoses, procedures, and pharmacy data. Data to be utilized in this study includes health plan enrollment data (i.e., dates of enrollment and disenrollment), demographic data (i.e., age, sex), pharmacy data (i.e., antiarrhythmic medication dispensings which are used to define the cohorts of interest), and diagnoses and procedure codes included in administrative claims data to assess cohort baseline characteristics (e.g., presence of atrial fibrillation, diabetes, coronary heart disease, heart failure). Diagnosis and procedure codes included in administrative claims data will also be used in combination with NDI cause of death information to define outcomes (e.g., diagnosis codes potentially associated with sudden cardiac death in claims data in conjunction with NDI cause of death information).

In this study, no new data are being collected. Instead, the study is being conducted using data from each health plan's database. Record level data including 18 HIPAA identifiers such as name, address (including city, county, or zip codes); contact information; social security, medical record, or health plan numbers; photos; or any other characteristic that could identify an individual will not be shared with the study coordinating center, or any other participating health plans. Instead, this information will be maintained behind each of the health plans firewalls, and will then send the required files directly to the NDI, receive matches directly from the NDI, and link matches to their data sets behind their institutional firewalls. No identifiers, including dates, will be returned to the coordinating center. Instead, the index date (i.e., date of antiarrhythmic medication dispensing) will be set to day "0". All other events, such as date of death, will be returned as the number of days since the date of dispensing or (e.g., "100"). Thus, the coordinating center will only receive dates as numbers – there will be no way to determine the exact date a patient began taking a medication, any other event.

Description of NDI+ data request: The study team will initiate a new request for death and cause of death data for study participants (antiarrhythmic users as well as a matched baseline cohort from the general population) at six health plans. [Lead Study Site] will coordinate the NDI+ linkage process by working with each of the participating health plans to assist with preparing files in accordance with NDI+ requirements.

7. EXTERNAL organizations (other than the NDI applicant's organization) receiving IDENTIFYING or IDENTIFIABLE death record information. (If there are no such external parties, type "None.")

[Definition of "identifying or identifiable death record information"—Any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Example: by identifying or identifiable death record data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identification numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual. (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)]

List the names of all EXTERNAL parties (organizations or outside consultants) who will obtain identifying or identifiable death record information from the NDI, from state vital statistics offices and/or from death record followback investigations. Include *administrative relationships* such as consultants, outside nosologists, contractors, subcontractors, sponsoring or participating agencies or organizations, and other major divisions or departments in your organization. If applicable, REPEAT some or all of the organizations listed in Items 3, 5 and 6 above. **(Note: Each organization listed in item 7 must complete and sign a Supplemental Confidentiality Agreement at the end of this application form.)**

IMPORTANT: Under each organization (or consultant) listed below, specify that organization's role and what project activities will be performed. Also specify (1) what identifying or identifiable death record information will be received, (2) in what form it will be received (e.g., death certificates or computer files), and (3) how the information will "flow" from one organization to another.

Names of Organizations and PI/Project Director*	Administrative Relationship (consultant, contractor, etc.)
A. [List Study Site#1] Site PI: [Name] Study Site #1 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #1)
B. [List Study Site#2] Site PI: [Name] Study Site #2 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #2)
C. [List Study Site#3] Site PI: [Name] Study Site #3 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #3)
D. [List Study Site#4] Site PI: [Name] Study Site #4 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #4)
E. [List Study Site#5] Site PI: [Name] Study Site #5 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #5)
F. [List Study Site#6] Site PI: [Name] Study Site #6 is requesting data from NDI in this application (cause of death).	[List Lead Study Site]holds a subcontract with institution. Co-Investigator (lead PI at Study Site #6)

*NOTE: A *"National Death Index Supplemental Confidentiality Agreement"* (see end of application form) must be completed by, or on behalf of, **each** organization (or individual) listed in items 3, 5 and 7 above and must be signed by responsible officials of that organization. This requirement is waived only for a FEDERAL GRANT listed in item 5 and then only when the NDI applicant gives assurances that identifying information obtained directly or indirectly from the NDI will not be provided to the granting agency.

8. Summary of Study Protocol or Project Activities

In responding to the following questions, please provide sufficient detail to describe your study or project and how data obtained via the NDI data will be used. Do not limit your responses to the space provided.
NOTE: Attach additional page if necessary.

8.a. Will the information obtained via NDI be included in a registry or any other type of study with long-term use and an indefinite end date? ☐ Yes ☒ No

What study type is this? (e.g., disease registry, longitudinal cohort study, cross-sectional study; case-control study)

Retrospective observational cohort study

All applicants must complete items 8.b and 8.c. If your application involves a registry, be sure also to include the following information in item 8.b. below: (1) the date the registry was founded, (2) the purpose of the registry, and (3) the eligibility criteria for including persons in the registry. A registry should also refer to Attachment B at the end of this application form for additional information to be included in item 8.c. below.

8.b. Purpose of study or project -- Describe the health or medical problem(s) addressed by your study or project. Include some background information to support why the study or project is being done. What are the primary objectives? If appropriate, include a description of hypotheses to be tested.

Certain medications may increase the risk of death or death from specific causes (e.g., sudden cardiac death[SCD]), but risks may not be identified in pre-market randomized trials. Capacity to examine death in post-market drug safety surveillance activities is essential to the U.S. Food and Drug Administration (FDA)'s mission to protect public health. While certain drugs, including antiarrhythmic medications, are known or suspected to increase the risk of death, FDA currently has limited ability to systematically examine mortality within its medical product safety surveillance system. Many health plan organizations have limited access to complete death or cause of death information required for FDA to systematically examine mortality. The NDI is currently the only complete national source of death and cause of death information accessible to population-based epidemiologic studies in the United States. Thus, linking multiple health plans databases included in this safety surveillance system to NDI+ data is critical for conducting population-based, post-market safety and effectiveness assessments of medical products with death and specific causes of death as outcomes.

The overall primary objective of this IRB approved study is to develop reusable administrative and technical processes for linking multiple health plan databases with NDI data so that the FDA may assess death and specific causes of death as outcomes in medical product safety and effectiveness studies in large distributed networks of electronic health plan databases. [Lead Study Site] as the lead site and coordinating center for the study will develop a distributed process for linkage that allows health plans to work directly with the NDI to eliminate sharing of identifiable patient information between participating health plans or with the coordinating center. The study team will pilot the developed approach through a use case designed to examine the associations between select antiarrhythmic medications and all-cause mortality and specific causes of death. Secondary objectives focus on utilizing linked health plan and NDI+ data to estimate rates of mortality and specific causes of death within the use case, and comparing them to rates previously reported in the literature. The outcomes of primary interest in the use case are all-cause mortality and SCD, but cardiovascular death (CV death) may also be examined if it is determined to be feasible by the study team.

The use case (select antiarrhythmic medications and their associations with all-cause mortality and specific causes of death) will guide the development and testing of these processes. We require NDI data to determine rates of mortality and specific causes of death (SCD, cardiovascular death) in a cohort of antiarrhythmic medication users and general population cohort.

NOTE: Attach additional page if necessary.

8.c. Study protocol or project activities—Please describe the study design, study population and sources of data, and study duration. Conclude your summary by describing how data obtained from the NDI, state death certificates, and death record "followback" investigations will be used. More information about death record followback investigations is requested in item 9. (NOTE: *Registries or long-term studies* should also refer to *Attachment B* at the end of this application form for additional information to be included in the response to item 8.c.)

See insert on next page.

NOTE: Attach additional page if necessary.

8.c. study protocol or project activities

Study design and background: This study will use a retrospective observational cohort study to examine mortality and specific causes of death in a cohort of antiarrhythmic medication users and a matched general population cohort. The study will develop a reusable process and workflow for distributed data networks to efficiently link multiple health plan databases to NDI+ data, with the goal of facilitating post-market safety and effectiveness assessments of medical products with death and specific causes of death as outcomes. Six data partners (aka health plans) have joined this NDI+ linkage study. Each health plan has used its electronic health information to develop a database that can be accessed to conduct research. These databases are formatted into a standard structure which uses standard variable names and coded values. Because all the databases are formatted in the same manner, the study coordinating center at [Lead Study Site] can write one program which is then run by each data partner against its local Sentinel database, returning deidentified individual-level or aggregated results to the coordinating center via a secure system. In this study, no identifiers, including dates, will be returned to the coordinating center. Instead, the index date (i.e., date of antiarrhythmic medication dispensing) will be set to day "0". All other events, such as date of death, will be returned as a number of days since the date of dispensing or (e.g., "100"). Thus, the coordinating center will only receive dates as numbers – there will be no way to determine the exact date a patient began taking a medication or any other event. This system protects the privacy and confidentiality of individual health information.

Sources of data: Electronic health information included in the distributed database (i.e., formatted into Common Data Model) maintained by each participating health plan, and death and cause of death information requested from NDI in this application. [Please note: NDI data are being requested just for the six sites listed in this application]. Cohorts will be identified at the health plans using data already available to them, including health plan data and electronic health records. Health plan data from the years 2000-ongoing (depending on NDI application approval timing, and health plan and NDI data availability) will be augmented by linkage to NDI+ data.

Study population: The study examines patients who initiate antiarrhythmic medications after 2000, as well as general population cohort matched to the antiarrhythmic medication cohort by age, health plan, and sex

Eligibility criteria:

*Ages 45+ years (45+ for men and 55+ women)
*Use of 1+ antiarrhythmic medications (oral dosage form of following medications: amiodarone, disopyramide, dofetilide, dronedarone, flecainide, mexiletine, propafenone, procainamide, quinidine, or sotalol) during the study period (antiarrhythmic medications cohort). National Drug Codes (NDCs) will be used to identify antiarrhythmic medication dispensing.
*Enrollment in medical and pharmacy benefits for 365 days prior to index date and no fills for any antiarrhythmic medications in prior 365 days (both cohorts)
*The general population cohort will be matched to the antiarrhythmic medication cohort by age and sex, also require enrollment in medical and pharmacy benefits for 365 days.

Study Duration: We are following patients from cohort entry-defining antiarrhythmic medication dispensing until end of use or loss to follow up (potential death, disenrollment from the health plan, end of database time). The general population cohort will be followed for the same timeframe as the antiarrhythmic medication user cohort.

How NDI data will be used: Death and cause of death data are being requested for cohort members within the six health plans included in this application. Data will be used to examine general all-cause and specific-cause mortality (sudden cardiac death, and cardiovascular death) in the general population and antiarrhythmic medications cohort. Data will be used to identify death and death from specific causes that occur in cohort members, and will augment health plan data generally, including any limited death information already available to the health plans. No death certificates will be requested using data from the NDI, and no "follow back" investigations will be conducted. This is a request for new records from NDI for all six health plans.

If approved for use, NDI death and cause of death data would be placed behind the fire wall at each health plan, health plans will run programs distributed by the coordinating center behind their firewalls, and provide ***aggregated results*** (e.g., counts of deaths, and specific causes of death in the cohorts) to the coordinating center. NDI data will remain at the health plans according to procedures described in this application and will not be made available for other purposes.

Should there be any significant deviations from the work described above, we fully understand that an amended NDI application must first be submitted to and approved by NCHS. All hard-copy death record information obtained via the NDI, will be flagged and stored separately from any administrative records or from statistical records that could be used in the future for purposes not described in the application. Computer records containing death record information obtained via the NDI shall also be flagged so that they will not be used in the future for purposes not described in the application. (Please see item 6 above for more information).

9. Death Record Follow-back Investigations

9.a. Does this study or project plan to perform "death record follow-back" investigations? [By "follow-back investigations" we mean that *once NDI identifies that certain study subjects are deceased*, your staff plans to collect additional information on those subjects' by going BACK to individuals or establishments that are (or would probably be) mentioned in the subjects' actual death certificates. This would include efforts to contact next-of-kin, physicians, hospitals and/or other parties appearing on the death certificates and/or already included in the decedents' research file.] NOTE: Follow-up refers to contacting the next-of-kin or health providers based on information already contained in researchers' file.

☐ Yes ☒ No (If yes, refer to *Attachment C* for additional documentation needed.)

9.b. If yes, what type of respondents will be contacted? Check all that apply.

☐ Decedent's next-of-kin

☐ Physicians

☐ Hospitals

☐ Other individuals or establishments mentioned on death record

9.c. What information will be obtained from EACH type of respondent?:

9.d. Name the organization(s) or consultant(s) who will be contacting EACH type of respondent:

9.e. Methods to be used in conducting followback investigations, including how EACH type of contact will be made:

10. Institutional Review Board (IRB) for the Protection of Human Subjects

(Defined by the U.S. Department of Health and Human Services in the Code of Federal Regulations, Title 45, Part 46)

Evidence of a current IRB review is REQUIRED for all NDI applications (please insure that applicant's name is referenced in the IRB letter). However, if this study or project involves death record "follow-back" investigations as described in item 9 above, a special letter from the IRB is REQUIRED (as explained in *Attachment C* at the end of the application form).

10.a. IRB approval status: ☐ Full ☒ Expedite ☐ Exempt

10.b. Attach a copy of the IRB review and provide the following:

Name of IRB: List Central IRB Name

IRB's Multiple Project Assurance (MPA) number or Federalwide Assurance (FWA) number: FWA00000xxx

Date of most current IRB review 12/18/2020

[NOTE: If death record "followback" investigations will be performed as described in item 9 above, an explanation of why your organization does not require an IRB approval for such a study or project is not acceptable. If your organization does not have an IRB (which has been approved by the Office for Human Research Protections, Department of Health and Human Services), you may have the study reviewed by an approved IRB in another organization.]

11. Obtaining State Death Certificates

11.a. Based on the results of the NDI file search(es), will copies of death certificates be requested from state vital statistics offices? ☐ Yes ☒ No

11.b. If you plan to request death certificates, what specific items of death certificate information do you expect to use in your analyses and/or to verify questionable matches? (Do not include NDI Plus variables; refer to NDI User's Guide for further information.)

12. Maintaining the Confidentiality of Identifying (or Identifiable) Information

12.a. Name the organization(s), including your own, which will:

(1) submit records of study subjects for the NDI file search(es):

A. Study Site #1 B. Study Site #2, C. Study Site #3, D. Study Site #4 E. Study Site #5, F. Study Site #6

(2) receive directly the results of the NDI search:

A. Study Site #1 B. Study Site #2, C. Study Site #3, D. Study Site #4 E. Study Site #5, F. Study Site #6

(3) request copies of death certificates from the state vital statistics offices:

N/A

12.b. Describe how your organization will store and maintain the confidentiality of the *identifying or identifiable death record information* obtained from (1) the NDI, (2) state death records, and (3) death record followback investigations. **"Identifying or identifiable death record information" refers to any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identification numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual.** (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

Describe the following controls that would be used to maintain the confidentiality of the NDI data:
NOTE: If multiple sites are involved in the above-mentioned study project, each site must describe its own controls that would be used to maintain the confidentiality of the NDI data.

- **Physical controls**–limiting access to the NDI data such as building guards, identification badges, key cards, closed circuit TV, and locked offices.
- **Technical controls** – such as user identification, passwords, firewalls, encryption, virtual private network, intrusion detection system, and stand-alone desktop use only.
- **Administrative controls** – such as how frequently files will be backed up, where backup files will be stored, methods in place to ensure least privilege access, methods for ensuring NDI identifying information is not co-mingled with administrative records not part of this project, how use of NDI data will be monitored to prevent its use for purposes other than those approved for this project, how personnel using the system will be trained and made aware of their responsibilities for protecting the NDI information, methods for keeping track of who has access to the data, and methods for ensuring return or destruction of data.

See insert on next page.

NOTE: Attach additional page if necessary.

12.b.

[Lead Study Site] - Study Coordinating Center

Physical Controls: [Lead Study Site] policies and procedures ensure controlled access to computers and physical space for secure storage of data and confidentiality information. Access to the [Lead Study Site] building is restricted by locked doors and requires a key card to enter [Lead Study Site] acility always. Reception staff supervises public entrances during normal operating hours and a roster of persons authorized to enter the area is maintained by administrative personnel. Any paper files with identifiers are required to be kept in locked filing cabinets. Keys are only provided to study staff once they are approved by the facilities and project managers.

Technical Controls: Data at [Lead Study Site] are managed in accordance with the national standards established by the HIPAA Security Rule and in accordance with the Federal Information Security Management Act of 2002 (FISMA). Administrative, physical, and technical safeguards are employed to ensure the confidentiality, integrity, and security of electronic health information (45 CFR Part 160 and Subparts A and C of Part 164; 44 U.S.C. § 3541, et seq). All computers require passwords to access the network and the electronic mail system. Passwords must be changed often (timing is dictated by [Lead Study Site] IT). Access to the computer systems is restricted to [Lead Study Site] staff. All computers used in the analysis are located in locked departments, all laptops are secured by Kensington lock wrap metal cable and data files are password protected. Staff needing to access the [Lead Study Site] network off-site must first be approved to do so. To log in off-site, individuals need to use their network password, a self-created pin (separate from password), and a time-sensitive unique code (gemalto token). Files are backed up several times a day on an encrypted server.

Administrative Controls:
[Lead Study Site] Pre-employment background checks must be completed before individuals are granted access to information systems and identifiers. Upon hire, all employees sign agreements to maintain confidentiality of data and research information. As a condition of employment, all members of the [Lead Study Site] workforce must complete training in Health Insurance Portability and Accountability Act (HIPAA) and Collaborative Institutional Training Initiative (CITI) Human Subjects research training requirements.

Individuals are granted access to specific files and programs based on job role. Staff can only access the [Lead Study Site] server only on approval by the study project manager or principal investigator and must be added as a study team member in study IRB application. [Lead Study Site] will receive cause of death data from NDI. We have already alerted the analysis and programming team which data collection sites will be transmitting cause of death data obtained from NDI. Because these data are not identifiable, we do not plan to destroy them. Instead, they will be kept in our final study analytic dataset and used in analyses.

[Lead Study Site] will not have direct access to death record information sent by NDI to the Data Partners. Each Data Partner who has access to NDI+ data will store its source data files behind a secured firewall on computers with anti-virus software. The data will be accessed only by the approved project staff working on this project. All investigators and project staff receive IRB and HIPAA training. Direct identifier and birth and death dates will be removed from study data files as soon as possible in the data processing steps. Unique study-specific identifiers will be assigned to support accurate linkage of data on the same individuals across multiple data files. Each Data Partner will maintain a secure cross-walk file linking the study identifier to the personal identifiers. Collected data will be used only for this project. Published data will not contain any individual identifiers. Please see the individual Data Partner
"Supplementary Confidentiality Agreements" for additional detail on how patient-level NDI data will be protected at each participating site.

13. Data Disposition Plan

Some state vital statistics offices have expressed concern about indefinite retention of *"identifying or identifiable death record information"* that could be used in the future by other persons for other purposes.

[Definition of "identifying or identifiable death record information" -- Any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identification numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual. (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)]

Except for data stored in registries, or other approved long-term studies, all identifying or identifiable data received from the NDI must be removed from all research records at the conclusion of the study or within 5 years after receipt of the NDI data -- regardless of the data set in which the data are kept. This means that all identifiers or potentially identifiable data elements associated with cause of death codes must be removed from all analysis files unless there is no way to identify an individual decedent. This also means that any linked files (with crosswalks) are to be destroyed. (Note: Death certificates obtained directly from state offices may have to be shredded in less than 5 years depending on each state's requirements.)

While the NDI staff recognizes that some research studies can remain active for several years, each study is viewed to have a limited duration. At the completion of the study or within 5 years after receipt of the NDI data ALL identifying or identifiable information that came from the NDI match must be destroyed, regardless of storage medium, unless no possible link could be made to an individual. **Note: As long as there are no identifiers or linkage variables remaining in the analytic or public-use file(s), cause(s) of death codes may remain in such file(s).**

1. Based on the above requirements, when do you plan to dispose of all identifying or identifiable death record information you obtained from the NDI? (Give the proposed month and year of destruction – or state UNKNOWN if this is an open-ended or ongoing study that has no specific disposition plan at this time.)

   MM YYYY

2. Only complete items 2.a. and 2.b. if the above date is UNKNOWN or if the date is more than 5 years after the month and year that you submitted this NDI Application Form.

   a. Please provide a strong justification of why the data need to be retained beyond this 5-year period.

   All identifiable data received from NDI will be removed from all research records within 5-years after receipt of the NDI data as noted above in the first paragraph. The MM YYYY date provided above is the estimated 5-years following receipt of the data from NDI and clarifying that this date is not indexed off of the date of this new application submission of MM YYYY. All identifying or identifiable data received from NDI will be destroyed the sooner of 5 years from receipt of the data or JMM YYYY.

   b. It is to be understood that within 5 years of submitting your NDI Application Form you are responsible for either (1) requesting an extension or (2) certifying the NDI data have been returned to NCHS or destroyed. (See attachment A) The extension request or certification of data disposal must be submitted to NDI staff within 5 years – no later than the month and year stated in the box below.

14. Completion of Study or Project

14.a. Indicate the scheduled termination date for the study, or whether the study is ongoing or open-ended. MM/DD/YYYY

14.b. In what form (e.g., aggregate, statistical, report, etc.) and to whom (e.g., peer reviewed scientific journals, monographs) will the results of your study or activities be released? (NDI would appreciate courtesy copy of any publications that may result from the use of NDI data)

Aggregate results from the results of the study will be shared to FDA and ultimately the public. All data published will be in aggregate form only so that no person could potentially be identified. (In addition, the study coordinating center and FDA will receive only de-identified data from participating health plans).

14.c. Will study subjects be notified of study results? ☐ Yes ☒ No

If *Yes*, how will the subjects be notified?

15. Other Uses of the Data

REMINDER: NDI data may not be used for legal, administrative, or other actions which may directly affect particular individuals or establishments as a result of their specific identification. NDI data may not be re-released to others except as specified in item 7 of this agreement. Re-release means providing access to, copies of, or in any other manner sharing, *individual-level identifying information* obtained from the NDI or the state death certificates to persons or organizations not specified in this approved NDI application form. The prohibition of re-release applies to microdata as well as to data in other form (e.g., copies of certificates). Aggregated tabular data without individual-level identifying information from the NDI or state death certificates are not considered re-release so long as the tabular data are not so detailed as to permit identification of individuals.

Will the identifying information (obtained from NDI, from state vital statistics offices, and/or from death record followback investigations) be used either directly or indirectly for any study or project other than the one described in "Summary of Study Protocol or Project Activities"? (See item 8 above.)

☐ Yes ☒ No ☐ Maybe

If *Yes* or *Maybe*, briefly describe the other purpose(s) for which the data will be used. (NOTE: A separate application form must be submitted for each study or project which will be using identifying information obtained via the NDI.)

16. Types of Data to Be Submitted to NCHS

16.a. Each record which you submit will be searched against records in the NDI file ONLY if your record contains at least one of the following combinations of data items: (Check all that apply.)

☒ First and last name and month and year of birth

☒ First and last name and Social Security Number

☒ Social Security Number, month, day and year of birth, and sex

16.b. Which of the following NDI data set items will you be able to provide for the records you submit? You are encouraged to provide as many of these data items as possible. This will maximize the number of true matches that are generated and will assist you in assessing the quality of the matches that occur.

	On approximately what percentage of your records?
1. First name	99 %
2. Middle name	56 %
3. Last name	99 %
4. Father's surname	99 %
5. Social Security Number (SSN)	60 %
6. Month of birth	99 %
7. Day of birth	99 %
8. Year of birth	99 %
9. Sex	99 %
10. Race	0 %
11. Marital status	0 %
12. State of residence[1]	99 %
13. State of birth	0 %
14. Age at death (if known)[2]	99 %
15. State of death[2]	50 %
16. Date or year of death[2]	99 %
17. Date or year of last contact[3]	99 %

[1]This item refers to the last known state of residence. The item is useful in assessing the matching results.

[2]For users submitting records for KNOWN decedents, these items are useful in assessing the matching results.

[3]For users submitting records for subjects whose vital status is UNKNOWN and for whom different years of death need to be searched, providing the date or year of last contact is useful in assessing the matching results.

National Death Index Confidentiality Agreement

Study or Project Title:

A Reusable, Generalizable Method to Link Health Plan Data with the National Death Index Plus to Examine the Associations Between Medical Products and Death and Causes of Death

The undersigned hereby agrees to the following terms and conditions associated with this National Death Index (NDI) application and to the use of the information obtained from (1) the NDI, (2) from State death records, and (3) from death record followback investigations:

A. Except for persons or organizations specified in the approved NDI application form, no data will be published or released in any form to any party if a particular individual or establishment is identifiable. **ALL REQUESTS FOR IDENTIFIABLE DATA OBTAINED VIA THE NDI WILL BE REFERRED IMMEDIATELY TO NCHS.** In accordance with Section 308(d) of the Public Health Service Act, such identifiable data will specifically not be provided in response to a direct order from an official of any government agency, the Administration or Congress, nor in response to an order from a court of justice.

B. The identifying information will be used ONLY for statistical purposes in medical and health research.

C. The identifying information will not be used as a basis for legal, administrative, or other actions which may directly affect those particular individuals or establishments as a result of their specific identification in this project.

D. The identifying information will be used only for the study or project proposed and the purpose described in the approved NDI application form. Use of the information for a research project other than the one described in the application form will not be undertaken until after a separate NDI application form for that project has been submitted to, and approved by, the NCHS.

E. NCHS obtains death record information via contracts with the state vital statistics offices. These contracts contain specific restrictions on the use of the information by the NDI and by the NDI Plus service (which gives NDI users cause of death codes). By providing NCHS with these assurances, I understand that I am also providing the same assurances to the state vital statistics offices. Violation of the terms and conditions of this Agreement may subject the organization/researcher to immediate abrogation of the Agreement by NCHS, the requirement of the return of all NDI data and related materials, and denial of future use of the NDI. Violation of the terms of the Agreement may also be a violation of Federal criminal law under 18 U.S.C. Section 1001. NCHS will pursue all legal remedies in the event of unauthorized disclosure of identifiable information from NDI data. Violation of the terms of the Agreement are also subject to state legal remedies.

F. The original version of the NDI data must be retained at a single location and no copy or extract of identifiable information may be made available to anyone except those persons identified in the Agreement and who have signed a non-disclosure statement. The NDI data may not be re-released to others except as specified in item seven of this agreement.

G. Access to identifiable NDI data maintained in computer memory must be controlled by password protection. Servers housing NDI data must be protected by a firewall and not be directly accessible from the internet. All persons must have completed required computer security training required by their institution. All printouts, diskettes, personal computers with data on hard disks, or other physical products containing identifiable information derived from the NDI must be kept in locked cabinets, file drawers, or other secure locations when not in use. Security procedures must be in place to ensure that identifiable NDI data cannot be used or taken by unauthorized individuals. Printouts, tabulations, reports and other materials must be edited for any possible disclosures of NDI identifiable data prior to making the information available to anyone other than those persons identified in this Agreement.

H. Except for data stored in registries or approved long term-studies, all identifying or identifiable data received from the NDI must be removed all research records at the conclusion of the study or within five years after receipt of the NDI data – regardless of the data set in which the data are kept—unless an extension has been granted by NDI. The original version of the NDI data must be returned to NCHS or destroyed. Files-- including backup files and derived files--with NDI identifying or identifiable data must be both deleted and overwritten to prevent recovery of the data. (See Attachment A.)

I. The organization/researcher must notify NCHS within 24 hours upon discovering any loss or suspected loss of identifiable NDI data or any disclosure of identifiable NDI data to unauthorized parties. This must be reported to the DVS Director, Delton Atkinson (919.541.2683). Within three business days of the notification to NCHS, the organization/researcher must submit to the NCHS Confidentiality Officer, a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

12

NDI Confidentiality Agreement (continued)

J. Authorized NCHS staff or agents may, upon request, be granted access to {name of user} facilities, where confidential NDI data are kept or used, for the purpose of inspecting the data security arrangements.

K. I understand that while state vital statistics offices may receive copies of this application, states may require additional information and/or assurances before responding to requests for copies of death certificates or for death record information. Some states may not be able to honor certain requests because of the proposed uses of the state data. Furthermore, once data from a particular state are received, I understand that users of the data are subject to that state's laws and regulations relating to disclosure of information on individuals or establishments.

L. I have reviewed this NDI application. All the statements made in this application and in any confidentiality assurances related to this application are true, complete, and correct to the best of my knowledge and belief. My signature below indicates my agreement to comply with the stated statutorily-based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates 18 USC 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

* ***NOTE***: The "official authorized to execute agreements" will vary among organizations. Whenever possible, the NDI prefers that this official be someone at a higher level of authority than the principal investigator or other persons responsible for the study or project; for example, a university official authorized to sign grant proposals, a company vice president, a government division or bureau director. By signing this agreement as the ***authorized official***, you are declaring that you have the authority to make the above assurances on behalf of the university, company, agency or other organization and to bind the organization to the terms of this agreement and you take responsibility for the confidentiality assurances of all organizations or individuals who are participating in this study.

For those individuals planning to sign digitally, please keep in mind that not all types of electronic signatures are acceptable. For further information see **Attachment D**.

The Data Steward for this project is: _____ Title: _____
Name

Organization: _____

Work phone number: _____ E-mail address: _____

As Data Steward, I affirm I will act as the custodian of the NDI files and will be responsible for the observance of conditions of use.

I will notify the NDI Director, Dr. Lillian Ingster (301-458-4286; LIngster@cdc.gov).

a. When access to the NDI data is no longer needed, (see Attachment
b. If a change in site access is contemplated;
c. Of the intent to modify the project's purpose; and
d. If these responsibilities are to be transferred.

Only complete this section if Lead Site is linking to the NDI

Signature of Data Steward: _____ Date: _____

SIGNATURE of *principal investigator or project director.*

Wet Signature Only

Signature Date

Name (Please type or print)

Title

Organization

E-mail: _____

***SIGNATURE** of *"official authorized to execute agreements" (last person to sign and date)*

Wet Signature Only

Signature Date

Name (Please type or print)

Title

Organization

E-mail: _____

Example of completed Supplemental Confidentiality form for Study Sites

National Death Index Supplemental Confidentiality Agreement

A separate Supplemental Confidentiality Agreement must be completed and signed by each *EXTERNAL* organization or consultant funding or participating in this study, as listed in *items 5 and 7* of the NDI Application Form. Co-Principal Investigators listed in *item 3* and employed in *external* organizations must also sign this Supplemental Confidentiality Agreement. The Supplemental Confidentiality Agreement(s) must then be submitted as an attachment to the Application Form. THIS REQUIREMENT IS WAIVED ONLY FOR A FEDERAL GRANT, AND THEN ONLY WHEN THE NDI APPLICANT (GRANTEE) CAN GIVE ASSURANCES THAT THE IDENTIFYING INFORMATION OBTAINED DIRECTLY OR INDIRECTLY FROM THE NDI WILL UNDER NO CIRCUMSTANCES BE PROVIDED TO THE GRANTOR.

Name and title of Principal Investigator, Project Director, Project Officer, or other responsible official:

Study Site #1 Principal Investigator (complete for each Study Site)

Organization name and complete mailing address:

Site A [list Study Site #1]
Street Address
City, State, Zip Code

Telephone Number: *E-mail :*

1. Will this organization (or individual) receive any of the identifying or identifiable death record information obtained from the NDI, state death records, and/or death record follow back investigations? (By "identifying or identifiable death record information" we mean any information on death certificates, other paper documents, or in computer files which by themselves, or if linked with other records, would permit the identification of one or more individuals or establishments. For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

☒ Yes ☐ No ☐ Maybe

2. Does this organization (or individual) have any contractual or other rights to the identifying information referred to above?

☐ Yes ☒ No ☐ Maybe

If you answered "No" to both questions 1 and 2, skip questions 3 and 4 below and just provide the two requested signatures below. If you answered "Yes" or "Maybe" to either questions 1 or 2, please complete questions 3 and 4 below and provide three signatures.

14

NDI Supplemental Confidentiality Agreement (continued)

3. In the box below, describe how your organization will store and maintain the confidentiality of the identifying or identifiable death record information obtained from (1) the NDI, (2) state death records, and (3) death record followback investigations. "Identifying or identifiable death record information" refers to any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identifica-tion numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual. (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

   Describe the following controls that would be used to maintain the confidentiality of the NDI data:

   - **Physical controls**–limiting access to data such as building guards, identification badges, key cards, closed circuit TV, and locked offices.
   - **Technical controls** – such as user identification, passwords, firewalls, encryption, virtual private network, intrusion detection system, and stand-alone desktop use only.
   - **Administrative controls** – such as how frequently files will be backed up, where backup files will be stored, methods in place to ensure least privilege access, methods for ensuring NDI identifying information is not co-mingled with administrative records not part of this project, how use of NDI data will be monitored to prevent its use for purposes other than those approved for this project, how personnel using the system will be trained and made aware of their responsibilities for protecting the NDI information, methods for keeping track of who has access to the data, and methods for ensuring return or destruction of data.

NOTE: If multiple sites are involved in the above-mentioned study project, each site must describe its own controls that would be used to maintain the confidentiality of the NDI data.

See insert below

4. **How and when will your organization dispose of identifying or identifiable death record data? If your organization has no plans to dispose of some or all of the identifying or identifiable death record data, please explain why.**

The study programmer at [list Study Site #1] will destroy files containing identifiable information by deleting crosswalk tables along with any other identifiable fields from NDI+ project files stored on our secure server. Original data received from NDI in CD/DVD media will be destoyed using a disk/DVD shredder. Original data that has been deleted cannot be restored. All identifying or identifiable data received from NDI will be destroyed the sooner of 5 years from receipt of the data or MM YYYY.

#3 Site: A. [List Study Site #1]

Physical Controls: Our Division of Pharmacoepidemiology computer system is physically located in a secure dedicated server room, managed by the [Study Site #1]. Policy 3.4 of the Data Core Policy informs the physical security of the data files, so that the following protections are followed: access to the server room is limited to the authorized systems administrator and project team members (when accompanied by the systems administrator).

Technical Controls: The project computer system operates entirely within the secured network, which is protected by a firewall maintained by the [Study Site #1] Information Technology Services, and access to data at our Division is controlled by accounts and passwords that comply with the requirements outlined in Policy 2.7, 2.8 and 2.9, as well as the appendix section of the Data Core Policy document: Electronic user identification codes limit access to specific directories and files based on roles and responsibilities. The Pharmacoepidemiology data server is behind the data center firewall. Remote access to the server is also restricted by a software firewall and IPSec filters. In addition to these physical and electronic security layers, all communications with the Pharmacoepidemiology server are encrypted.

Administrative Controls: NDI files received for this application will be backed up daily on encrypted backup disks which are stored in a locked cabinet accessible only by study team members. The data received from the NDI will be kept in a dedicated, separate NDI secure folder, separate from administrative datasets/records and will only be accessed by a study team member in the study IRB application. All investigators and project staff receive IRB, HIPAA and annual compliance training to ensure that all PHI information received from the NDI is protected. Individuals are granted access to specific files and programs based on job role. The data received from NDI will be accessible only by the study programmer using a remote desktop application and secured through the use of individually unique user identification codes and passwords. Furthermore, study staff can only access the NDI files via approval from Principal Investigator or Project Manager. Collected data will be used only for this IRB approved study and will not be combined with any non-study administrative datasets/records. At project completion, the NDI data will be archived with an automated, mandatory deletion date the sooner of 5 years from receipt of the data or July 2024. Published data will not contain any individual identifiers.

NDI Supplemental Confidentiality Agreement (continued)

Study or Project Title:

A Reusable, Generalizable Method to Link Health Plan Data with the National Death Index Plus to Examine the Associations Between Medical Products and Death and Causes of Death

5. The undersigned hereby agrees to the following terms and conditions associated with this National Death Index (NDI) application and to the use of the information obtained from (1) the NDI, (2) from State death records, and (3) from death record follow back investigations:

   A. Except for persons or organizations specified in the approved NDI application form, no data will be published or released in any form to any party if a particular individual or establishment is identifiable. ALL REQUESTS FOR IDENTIFIABLE DATA OBTAINED VIA THE NDI WILL BE REFERRED IMMEDIATELY TO NCHS. In accordance with Section 308(d) of the Public Health Service Act, such identifiable data will specifically not be provided in response to a direct order from an official of any government agency, the Administration or Congress, nor in response to an order from a court of justice.

   B. The identifying information will be used ONLY for statistical purposes in medical and health research.

   C. The identifying information will not be used as a basis for legal, administrative, or other actions which may directly affect those particular individuals or establishments as a result of their specific identification in this project.

   D. The identifying information will be used only for the study or project proposed and the purpose described in the approved NDI application form. Use of the information for a research project other than the one described in the application form will not be undertaken until after a separate NDI application form for that project has been submitted to, and approved by, the National Center for Health Statistics.

   E. NCHS obtains death record information via contracts with the state vital statistics offices. These contracts contain specific restrictions on the use of the information by the NDI and by the NDI Plus service (which gives NDI users cause of death codes). By providing NCHS with these assurances, I understand that I am also providing the same assurances to the state vital statistics offices. Violation of the terms and conditions of this Agreement may subject the organization/researcher to immediate abrogation of the Agreement by NCHS, the requirement of the return of all NDI data and related materials, and denial of future use of the NDI. Violation of the terms of the Agreement may also be a violation of Federal criminal law under 18 U.S.C. Section 1001. NCHS will pursue all legal remedies in the event of unauthorized disclosure of identifiable information from NDI data. Violation of the terms of the Agreement are also subject to state legal remedies.

   F. The original version of the NDI data must be retained at a single location and no copy or extract of identifiable information may be made available to anyone except those persons identified in the Agreement and who have signed a non-disclosure statement. The NDI data may not be re-released to others except as specified in item 7 of this agreement.

   G. Access to identifiable NDI data maintained in computer memory must be controlled by password protection. Servers housing NDI data must be protected by a firewall and not be directly accessible from the internet. All persons must have completed required computer security training required by their institution. All printouts, diskettes, personal computers with data on hard disks, or other physical products containing identifiable information derived from the NDI must be kept in locked cabinets, file drawers, or other secure locations when not in use. Security procedures must be in place to ensure that identifiable NDI data cannot be used or taken by unauthorized individuals. Printouts, tabulations, reports and other materials must be edited for any possible disclosures of NDI identifiable data prior to making the information available to anyone other than those persons identified in this Agreement.

   H. Except for data stored in registries or approved long term-studies, all identifying or identifiable data received from the NDI must be removed all research records at the conclusion of the study or within five years after receipt of the NDI data – regardless of the data set in which the data are kept—unless an extension has been granted by NDI. The original version of the NDI data must be returned to NCHS or destroyed. Files-- including backup files and derived files--with NDI identifying or identifiable data must be both deleted and overwritten to prevent recovery of the data. (See Attachment A.)

   I. The organization/researcher must notify NCHS within 24 hours upon discovering any loss or suspected loss of identifiable NDI data or any disclosure of identifiable NDI data to unauthorized parties. This must be reported to the DVS Director, Delton Atkinson (919.541.2683). Within 3 business days of the notification to NCHS, the organization/researcher must submit to the NCHS Confidentiality Officer, a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

16

NDI Supplemental Confidentiality Agreement (continued)

J. Authorized NCHS staff or agents may, upon request, be granted access to {name of user} facilities, where confidential NDI data are kept or used, for the purpose of inspecting the data security arrangements.

K. I understand that while state vital statistics offices may receive copies of this application, states may require additional information and/or assurances before responding to requests for copies of death certificates or for death record information. Some states may not be able to honor certain requests because of the proposed uses of the state data. Furthermore, once data from a particular state are received, I understand that users of the data are subject to that state's laws and regulations relating to disclosure of information on individuals or establishments.

L. I have reviewed this NDI application. All the statements made in this application and in any confidentiality assurances related to this application are true, complete, and correct to the best of my knowledge and belief. My signature below indicates my agreement to comply with the stated statutorily-based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates 18 USC 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

NOTE: If your response to both items 1 and 2 above was "No", you must still sign this form below; **HOWEVER**, it is understood that the terms specified in item 5 above do not apply to you or to your organization. And, because you will not be receiving identifiable NDI data, you would not need a Data Steward's signature.

* ***NOTE***: The "official authorized to execute agreements" will vary among organizations. Whenever possible, the NDI prefers that this official be someone at a higher level of authority than the principal investigator or other persons responsible for the study or project; for example, a university official authorized to sign grant proposals, a company vice president, a government division or bureau director. By signing this agreement as the ***authorized official***, you are declaring that you have the authority to make the above assurances on behalf of the university, company, agency or other organization and to bind the organization to the terms of this agreement and you take responsibility for the confidentiality assurances of all organizations or individuals who are participating in this study.

For those individuals planning to sign digitally, please keep in mind that not all types of electronic signatures are acceptable. For further information see **Attachment D**.

The Data Steward for this project is: Study Site Programmer (must sign first) Title: _____
Name

Organization: _____

Work phone number: _____ E-mail address: _____

As Data Steward, I affirm I will act as the custodian of the NDI files and will be responsible for the observance of conditions of use.

I will notify the NDI Director, Dr. Lillian Ingster (301-458-4286; LIngster@cdc.gov).

a. When access to the NDI data is no longer needed, (see Attachment A);
b. If a change in site access is contemplated;
c. Of the intent to modify the project's purpose; and
d. If these responsibilities are to be transferred.

Wet signatures only

Signature of Data Steward: _____ Date: _____

SIGNATURE of *Principal Investigator, Project Director, or Project Officer.*

Signature Date

Name (Please type or print)

must sign second

Title

Organization

E-mail: _____

***SIGNATURE** of *"official authorized to execute agreements" (last person to sign and date)*

Signature Date

Name (Please type or print)

must sign third

Title

Organization

E-mail: _____

17

Example of completed Supplemental Confidentiality form for Study Sponsor

National Death Index Supplemental Confidentiality Agreement

A separate Supplemental Confidentiality Agreement must be completed and signed by each *EXTERNAL* organization or consultant funding or participating in this study, as listed in *items 5 and 7* of the NDI Application Form. Co-Principal Investigators listed in *item 3* and employed in *external* organizations must also sign this Supplemental Confidentiality Agreement. The Supplemental Confidentiality Agreement(s) must then be submitted as an attachment to the Application Form. THIS REQUIREMENT IS WAIVED ONLY FOR A FEDERAL GRANT, AND THEN ONLY WHEN THE NDI APPLICANT (GRANTEE) CAN GIVE ASSURANCES THAT THE IDENTIFYING INFORMATION OBTAINED DIRECTLY OR INDIRECTLY FROM THE NDI WILL UNDER NO CIRCUMSTANCES BE PROVIDED TO THE GRANTOR.

Name and title of Principal Investigator, Project Director, Project Officer, or other responsible official: Study Sponsor Principal Investigator

Organization name and complete mailing address:

Telephone Number: *E-mail :*

1. Will this organization (or individual) receive any of the identifying or identifiable death record information obtained from the NDI, state death records, and/or death record follow back investigations? (By "identifying or identifiable death record information" we mean any information on death certificates, other paper documents, or in computer files which by themselves, or if linked with other records, would permit the identification of one or more individuals or establishments. For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

☐ Yes ☒ No ☐ Maybe

2. Does this organization (or individual) have any contractual or other rights to the identifying information referred to above?

☐ Yes ☒ No ☐ Maybe

If you answered "No" to both questions 1 and 2, skip questions 3 and 4 below and just provide the two requested signatures below. If you answered "Yes" or "Maybe" to either questions 1 or 2, please complete questions 3 and 4 below and provide three signatures.

14

NDI Supplemental Confidentiality Agreement (continued)

3. In the box below, describe how your organization will store and maintain the confidentiality of the identifying or identifiable death record information obtained from (1) the NDI, (2) state death records, and (3) death record followback investigations. "Identifying or identifiable death record information" refers to any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identifica-tion numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual. (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)

   Describe the following controls that would be used to maintain the confidentiality of the NDI data:

   - **Physical controls**–limiting access to data such as building guards, identification badges, key cards, closed circuit TV, and locked offices.
   - **Technical controls** – such as user identification, passwords, firewalls, encryption, virtual private network, intrusion detection system, and stand-alone desktop use only.
   - **Administrative controls** – such as how frequently files will be backed up, where backup files will be stored, methods in place to ensure least privilege access, methods for ensuring NDI identifying information is not co-mingled with administrative records not part of this project, how use of NDI data will be monitored to prevent its use for purposes other than those approved for this project, how personnel using the system will be trained and made aware of their responsibilities for protecting the NDI information, methods for keeping track of who has access to the data, and methods for ensuring return or destruction of data.

NOTE: If multiple sites are involved in the above-mentioned study project, each site must describe its own controls that would be used to maintain the confidentiality of the NDI data.

Section does not need to be completed

4. **How and when will your organization dispose of identifying or identifiable death record data? If your organization has no plans to dispose of some or all of the identifying or identifiable death record data, please explain why.**

Section does not need to be completed

NDI Supplemental Confidentiality Agreement (continued)

Study or Project Title:

A Reusable, Generalizable Method to Link Health Plan Data with the National Death Index Plus to Examine the Associations Between Medical Products and Death and Causes of Death

5. The undersigned hereby agrees to the following terms and conditions associated with this National Death Index (NDI) application and to the use of the information obtained from (1) the NDI, (2) from State death records, and (3) from death record follow back investigations:

   A. Except for persons or organizations specified in the approved NDI application form, no data will be published or released in any form to any party if a particular individual or establishment is identifiable. ALL REQUESTS FOR IDENTIFIABLE DATA OBTAINED VIA THE NDI WILL BE REFERRED IMMEDIATELY TO NCHS. In accordance with Section 308(d) of the Public Health Service Act, such identifiable data will specifically not be provided in response to a direct order from an official of any government agency, the Administration or Congress, nor in response to an order from a court of justice.

   B. The identifying information will be used ONLY for statistical purposes in medical and health research.

   C. The identifying information will not be used as a basis for legal, administrative, or other actions which may directly affect those particular individuals or establishments as a result of their specific identification in this project.

   D. The identifying information will be used only for the study or project proposed and the purpose described in the approved NDI application form. Use of the information for a research project other than the one described in the application form will not be undertaken until after a separate NDI application form for that project has been submitted to, and approved by, the National Center for Health Statistics.

   E. NCHS obtains death record information via contracts with the state vital statistics offices. These contracts contain specific restrictions on the use of the information by the NDI and by the NDI Plus service (which gives NDI users cause of death codes). By providing NCHS with these assurances, I understand that I am also providing the same assurances to the state vital statistics offices. Violation of the terms and conditions of this Agreement may subject the organization/researcher to immediate abrogation of the Agreement by NCHS, the requirement of the return of all NDI data and related materials, and denial of future use of the NDI. Violation of the terms of the Agreement may also be a violation of Federal criminal law under 18 U.S.C. Section 1001. NCHS will pursue all legal remedies in the event of unauthorized disclosure of identifiable information from NDI data. Violation of the terms of the Agreement are also subject to state legal remedies.

   F. The original version of the NDI data must be retained at a single location and no copy or extract of identifiable information may be made available to anyone except those persons identified in the Agreement and who have signed a non-disclosure statement. The NDI data may not be re-released to others except as specified in item 7 of this agreement.

   G. Access to identifiable NDI data maintained in computer memory must be controlled by password protection. Servers housing NDI data must be protected by a firewall and not be directly accessible from the internet. All persons must have completed required computer security training required by their institution. All printouts, diskettes, personal computers with data on hard disks, or other physical products containing identifiable information derived from the NDI must be kept in locked cabinets, file drawers, or other secure locations when not in use. Security procedures must be in place to ensure that identifiable NDI data cannot be used or taken by unauthorized individuals. Printouts, tabulations, reports and other materials must be edited for any possible disclosures of NDI identifiable data prior to making the information available to anyone other than those persons identified in this Agreement.

   H. Except for data stored in registries or approved long term-studies, all identifying or identifiable data received from the NDI must be removed all research records at the conclusion of the study or within five years after receipt of the NDI data – regardless of the data set in which the data are kept—unless an extension has been granted by NDI. The original version of the NDI data must be returned to NCHS or destroyed. Files-- including backup files and derived files--with NDI identifying or identifiable data must be both deleted and overwritten to prevent recovery of the data. (See Attachment A.)

   I. The organization/researcher must notify NCHS within 24 hours upon discovering any loss or suspected loss of identifiable NDI data or any disclosure of identifiable NDI data to unauthorized parties. This must be reported to the DVS Director, Delton Atkinson (919.541.2683). Within 3 business days of the notification to NCHS, the organization/researcher must submit to the NCHS Confidentiality Officer, a more detailed written report including the date and nature of the event, actions taken or to be taken to remediate the issue(s), and plans or processes developed to prevent further problems, including specific information on timelines anticipated for action.

16

NDI Confidentiality Agreement (continued)

J. Authorized NCHS staff or agents may, upon request, be granted access to {name of user} facilities, where confidential NDI data are kept or used, for the purpose of inspecting the data security arrangements.

K. I understand that while state vital statistics offices may receive copies of this application, states may require additional information and/or assurances before responding to requests for copies of death certificates or for death record information. Some states may not be able to honor certain requests because of the proposed uses of the state data. Furthermore, once data from a particular state are received, I understand that users of the data are subject to that state's laws and regulations relating to disclosure of information on individuals or establishments.

L. I have reviewed this NDI application. All the statements made in this application and in any confidentiality assurances related to this application are true, complete, and correct to the best of my knowledge and belief. My signature below indicates my agreement to comply with the stated statutorily-based requirements with the knowledge that deliberately making a false statement in any matter within the jurisdiction of any department or agency of the Federal Government violates 18 USC 1001 and is punishable by a fine of up to $10,000 or up to 5 years in prison.

* ***NOTE***: The "official authorized to execute agreements" will vary among organizations. Whenever possible, the NDI prefers that this official be someone at a higher level of authority than the principal investigator or other persons responsible for the study or project; for example, a university official authorized to sign grant proposals, a company vice president, a government division or bureau director. By signing this agreement as the ***authorized official***, you are declaring that you have the authority to make the above assurances on behalf of the university, company, agency or other organization and to bind the organization to the terms of this agreement and you take responsibility for the confidentiality assurances of all organizations or individuals who are participating in this study.

For those individuals planning to sign digitally, please keep in mind that not all types of electronic signatures are acceptable. For further information see **Attachment D**.

The Data Steward for this project is: _____ Title: _____
Name

Organization: _____

Work phone number: _____ E-mail address: _____

As Data Steward, I affirm I will act as the custodian of the NDI files and will be responsible for the observance of conditions of use.
I will notify the NDI Director, Dr. Lillian Ingster (301-458-4286; LIngster@cdc.gov).

a. When access to the NDI data is no longer needed, (see Attachment
b. If a change in site access is contemplated;
c. Of the intent to modify the project's purpose; and
d. If these responsibilities are to be transferred.

This section does not need to be completed by the Study Sponsor

Signature of Data Steward: _____ Date: _____

SIGNATURE of *principal investigator or project director.*	*SIGNATURE of "*official authorized to execute agreements" (last person to sign and date)*
Signature Date	Signature Date
Name (Please type or print)	Name (Please type or print)
Title	Title
Organization	Organization
E-mail:	E-mail:

NDI may accept electronic signatures for the Study Sponsor. It is advised to check with the NDI advisors.

13

VIII. NDI Data Disposition Form

Attachment A

National Death Index (NDI) Data Disposition Form

Use the multi-purpose form on the next page to notify the NDI program of one of the following events:

- When you have disposed of ALL the identifying or identifiable death record information obtained from the NDI
- If your initial NDI Application was submitted more than 5 years ago and you are now submitting a NDI Repeat Request Form (and have never completed this form)
- To request an extension for the retention of your identifying or identifiable death record information beyond 5 years from when your initial NDI application was submitted.
- If you have already been approved for a 1 to 5 year extension, to request another extension beyond your previously approved extension period.

Some state vital statistics offices have expressed concern about indefinite retention of *"identifying or identifiable death record data"* that could be used in the future by other persons for other purposes.

[Definition of "identifying or identifiable death record data" -- Any information on death certificates, other paper documents, or in computer files which by itself, or if linked with other records, would permit the identification of one or more individuals or establishments. Furthermore, by identifying or identifiable data we mean such items as name(s), Social Security Number, exact dates, addresses, and death certificate number. Even with the removal of direct identifiers and linkable study subject identification numbers, there is still a special concern that some combinations of the remaining variables could potentially be used to identify an individual. (For example: a combination of date-of-birth, date-of-death and/or cause-of-death is considered identifiable)]

Except for data stored in registries or other approved long term studies, all identifying or identifiable death record data received from the NDI must be removed from all research records at the conclusion of the study or within 5 years after receipt submission of your initial NDI Application Form -- regardless of the data set in which the data are kept. This means that all identifiers or potentially identifiable data elements associated with cause of death codes must be removed from all analysis files unless there is no way to identify an individual decedent. This also means that any linked files (with crosswalks) are to be destroyed. (Note: Death certificates obtained directly from state offices may have to be shredded in less than 5 years depending on each state's requirements.)

While the NDI staff recognizes that some research studies can remain active for several years, each study is viewed to have a limited duration. At the completion of the study ***all*** identifying or identifiable death record data that came from the NDI match must be destroyed, regardless of storage medium, unless no possible link could be made to an individual. **Note: As long as there are no identifiers or linking variables remaining in the analytic or public-use file(s), cause(s) of death codes may remain in such file(s).**

Attachment A

NDI Data Disposition Form (continued)

Date Request Approved	NDI Application Number

Title of study or project:

Principal Investigator or Project Director:

Title:

Organization

Mailing address:

Phone no.: E-mail:

1. As the Data Custodian for the above listed study/project, I affirm that all electronic and paper files containing identifiable NDI data have been destroyed on: (If not destroyed, put NA and answer items 3 – 5 below.)
2. I also affirm that all derivative and back-up copies have been destroyed on: (If not destroyed yet, put NA and answer items 3 – 5 below.)
3. When will the identifiable death record information be destroyed? (State UNKNOWN if this is an open-ended or ongoing study that has no specific disposition plan at this time.)
4. If the answer to item 3 is: (1) unknown, (2) more than 5 years after you submitted your NDI Application Form, or (3) more than 5 years after you last requested an extension for the retention of your data, please provide a strong justification for why the data need to be retained beyond the 5-year period.
5. If it has been more than 5 years since your initial NDI application (or since your last request for an extension), are you requesting an extension for the retention of identifiable NDI data? ☐ Yes ☐ No
6. If your extension is approved, you are responsible for submitting this form when your data have been destroyed OR within 5 years from now but no later than the date you indicate in the box to the right.

Data Steward (print name and title)	Signature	Date

Principal Investigator or Project Director (print name and title)	Signature	Date

Mail form to: National Death Index, NCHS, 3311 Toledo Road, Room 5292, Hyattsville, MD 20782

12

IX. Lessons Learned from Piloting a Central IRB Process and Submitting an NDI Application

This section summarizes challenges and recommendations to assist future studies with efficiently achieving IRB approval for multi-site studies which require linkage to data from the NDI.

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
1	IRB Cede process	State or local laws or policies for using death data	Study Sites may be in states with more restrictive laws concerning use of death data and may have implications on requirements for IRB review. We learned that Study Sites may need to solicit state approval for using death data prior to IRB submission.	At project initiation, determine requirements regarding any state or local laws or policies for using death data at Study Sites before the initial IRB submission.	**1**

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
2	IRB Cede Process	Study Site data sources and governance	Health Plans with Medicaid contracts for the acquisition of the NDI data. These Health Plans may be required to submit an IRB application to the State IRB to ensure compliance with the U.S. Department of Health and Human Services regulations for the protection of human subjects. This resulted in minor delays for soliciting IRB approval as the State IRB convene monthly.	At the onset of study initiation, confirm with all Studies Sites if there are additional IRB approvals required for additional acquired data sources (i.e., Medicaid data) and if the approvals can be done simultaneously or not.	1

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
3	Initial IRB Submission	NDI data disposition requirements	NDI in most cases requires that NDI data must be disposed within five years of the completion of the NDI application (unless exception granted/appropriate paperwork filed). Study Sites may follow different data disposition requirements and need to ensure that sites are aware of NDI's data retention policy.	Ensure that the Central IRB and Study Sites are aware of the NDI data disposition and data protection requirements and any study agreements with the NDI. This information should be included in the HIPAA waiver, if a waiver applies. The IRB of record also reviews detail on how patient-level NDI data will be protected and destroyed at each Study Site.	1
4	NDI application process	Draft NDI application submission	The Coordinating Center did not initially submit draft legal agreements for NDI review.	The NDI provided feedback that draft legal agreements should be provided along with the draft application to allow them to provide feedback on essential language regarding NDI security and data retention policies.	1

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
5	NDI application process	NDI application: data disposition policy	The NDI advisors that NDI data disposition policy is driven by the date of the final application submission, not date of NDI data receipt. This is important as it will inform how legal agreements and the NDI application is completed.	Study teams should plan to destroy data or file an extension with the NDI within 5 years of application submission date. The Coordinating Center should confirm data destruction plans at the conclusion of the study and designate a point of contact (preferably the Site Principal Investigator) to confirm the NDI Data Disposition Form has been completed.	1
6	IRB Cede process	Drafting one HIPAA waiver form (if applicable)	Each Study Site drafted their own HIPAA waiver form for the project. The Coordinating Center learned that one HIPAA waiver should be drafted and reviewed by Study Sites before submission to the lead IRB.	The Coordinating Center should draft the HIPAA waiver and distribute to Study Sites for review and approval prior to submission to the lead IRB.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
7	Initial IRB Submission	Changes to study population following submission of the study proposal to the IRB	Study population changes occurring between a proposal used to open an initial IRB application and final protocol may results in delays in the timeline for IRB approval. A technical proposal was used to open the study with the overseeing IRB. However, the use case was modified for the final protocol in a way that changed the study population. This study population modification required an amendment to the initial IRB study application, additional review by the Study Sites, as well as a feasibility analysis to support protocol development. This created some minor inefficiencies in the IRB review process.	A couple of Study Sites required a final protocol review before agreeing to cede review to a Central IRB (i.e., draft protocol was not sufficient). This is something future studies should account for in their study timelines as needed.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
8	IRB Cede Process	Study Site specific Information Technology (IT) security requirements	One Study Site IRB required a technical risk assessment evaluation before proceeding with approval to cede to the Central IRB. The risk assessment involved the evaluation of the proposed Central IRB's assurance of protection for PHI data that may be exchanged between the Study Site and Coordinating Center. This required an extra step in the cede process, and delayed soliciting cede approval for this site.	At the onset of study initiation, the Study Coordinating Center should perform requirements gathering regarding any IT security requirements to allow data exchange with the Coordinating Center or the NDI.	2
9	NDI Application	Data Use Agreements (DUA)	One Study Site inquired if NDI has a federal certificate of confidentiality, and worked with the NDI to set up a data use agreement (DUA). \	At the onset of project, confirm with Study Sites if they require a DUA to exchange data with NDI to complete NDI linkage.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
10	Master Reliance Agreements (MRAs) Process	Policies and Procedures for MRAs	While Master Reliance Agreements may be leveraged for efficiency in the Central IRB review process, Study Sites may be members of separate research networks that follow different policies & procedures, including platforms for documenting reliance. This may require extra review steps the lead IRB and create timeline delays.	To establish an efficient Central IRB process, discuss options for IRB Master Reliance Agreements with Study Sites at the onset of the study. These discussions will assist the Coordinating Center with navigating policies and procedures for any respective MRAs that need to be accommodated within the study timeline	3
11	NDI Application	Submission timeline	The NDI application submission volume is typically higher in April-May and August-September. Submissions around this timeframe may have a longer review period given the volume of submissions that need NDI Board review.	Study teams should plan their application submission accordingly and avoid high-volume submission periods if possible, to avoid a lengthy application review period.	3

Attachment 2: Supplemental Appendix: Distributed Program Specifications Implemented in a Pilot Study

I. Pilot Project Cohort Specifications for Use Case Example

Use case specifications

Inclusion and exclusion criteria for the use case
Our pilot project utilized data captured within participating health plan databases between 2000 and 2017 (or earliest or latest available health plan data, and most recent NDI data available at time of NDI application), augmented by linkage to NDI data.

Cohort 1 included new users of selected antiarrhythmic medications **(Table B1)** for males aged 45 years and above and females aged 55 years and above on the date of cohort entry. The list of select antiarrhythmic medications of interest and new-user definition is described under the "Exposure identification for the use case" section below**.**

For all cause-mortality and cardiovascular (CV) death analyses, the entire cohort was used. For analyses focused on sudden cardiac death (SCD), cohorts was restricted to individuals under the age of 75 years to maintain consistency with Chung et al.[37] Chung et al removed patients aged 75 years or older to minimize false positives as the risk of mortality increases with age, and the study also found death certificates to be less reliable for identifying SCD in older individuals. Although it may be difficult to capture nursing home stays within health plan databases, to maintain consistency with the Chung et al algorithm, individuals with evidence of a nursing home stay in the baseline period were excluded. Cohort 1 entry began upon an individual's first dispensing for an oral dosage form of an antiarrhythmic medication of interest and:

a. preceded by a 365-day baseline period with medical and pharmacy benefits (gaps in enrollment <45 days ignored); and
b. preceded by a 365-day baseline period in which the individual has ≥1 encounter with a diagnosis or medication dispensing

To approximate typical drug safety study situations in which no future information was used to determine medication users' vital status, individuals with more than one episode of new use during the period of interest only contributed their first episode. This study design choice helped avoid the selection bias that use of future information may generate. The protocol allowed gaps in enrollment of <45 days because it is believed that these do not represent true gaps in coverage but instead administrative changes. Index date would be the date of the first eligible dispensing for the select antiarrhythmic drugs of interest. Treatment episodes were created based on dispensings, and the end date of treatment was calculated based on dispensing information.

Cohort 2 (i.e., average-risk cohort) was drawn from individuals who were not current (i.e., on day of cohort entry) or past (i.e., prior 365 days) users of antiarrhythmic medications of interest **(Table B1).** Among such individuals, Cohort 2 was matched at a ratio of one-to-one to Cohort 1 based on age, sex, and health plan. Index dates were matched to Cohort 1.

Individuals in Cohort 2 were required to have a 365-day baseline period with medical and pharmacy benefits (gaps in enrollment <45 days ignored as specified above in Cohort 1) and at least one medical encounter or pharmacy dispensing claim in the prior 365 days. As in Cohort 1, Cohort 2 included the entire cohort for all cause-mortality and CV death analyses but were restricted to individuals under the age of 75 years and no evidence of a nursing home stay in the baseline period for SCD analyses.

Different age cut offs for men and women were selected as risks of all-cause mortality and SCD vary considerably by sex. The goal was to improve the specificity of mortality and specific causes of death outcomes identified through NDI+ data matching. Younger individuals are less likely to experience mortality and SCD than older individuals, and within age groups, women are less likely to experience mortality and SCD than men. The risk for SCD has been shown to increase in women after age 55 years.[38] All-cause mortality is also rare in younger age groups. Thus, if we were to identify women under 55 years who appeared to have disenrolled from the health plan and send them to NDI for matching, a larger proportion would represent true health plan disenrollment (without death) than for men in the same age range. Choosing a higher age cut-off for women is intended to decrease "false positive" matches and minimize the number of NDI submissions.

Exposure identification for the use case

Selected oral antiarrhythmic medications of interest are included in **Table A1** and was identified with National Drug Codes (NDCs). New use will be defined by excluding individuals with dispensings of class I and III antiarrhythmic drugs (all routes of administration), including amiodarone, disopyramide, dofetilide, dronedarone, encainide, flecainide, ibutilide, mexiletine, moricizine, procainamide, propafenone, quinidine, sotalol, and tocainide,[39,40] in the 365-day baseline period. Individuals with dispensings of Lidocane (intravenous only) in the 365-day baseline period were excluded. Baseline exposure to adenosine A1 agonists, digoxin, phenytoin, class II β-blocker agents, and class IV (i.e., calcium channel blocker) agents were ignored, as these drugs/classes are not used specifically for the treatment of arrhythmias.

When creating treatment episodes, we applied a stockpiling algorithm that accounted for the fact that members might refill prescriptions before the end of days supply of their prior prescription. For example, if a member received a 30-day dispensing for sotalol on January 1st and then received a second 30-day dispensing on January 20th, the stockpiling algorithm will adjust the second dispensing so that it starts on January 31st, after the first dispensing has been used in full. The treatment episode would thus be 60 days in total, through March 1st (assuming February has 28 days). We also implemented a 14-day episode gap when creating treatment episodes to account for members with imperfect adherence. An episode gap was the maximum number of days of interrupted days-supply allowed between two claims for the same drugs of interest. If the number of days between when one prescription claim ran out and the next claim was smaller than or equal to the episode gap, the algorithm "bridged" these two claims to build a continuous treatment episode. If, however, the number of days between the two claims of the same treatment exceeded the episode gap, the treatment episode ended at the end of the 14-day period. The episode gap was assessed after claim service dates were adjusted by the stockpiling algorithm. Because we were interested in the risk of all-cause mortality and SCD for the class of medications in general and not individual antiarrhythmic medications, analyses would focus on users of any antiarrhythmic medications of interest as a group, and results would not be stratified by individual medication.

Individuals only contributed their first episode of antiarrhythmic medication use included in health plan databases. Despite best efforts to restrict the cohorts, individuals included in either Cohort 1 or 2 might in fact have used antiarrhythmics medications in the past (e.g., outside of the study period or before enrolling in a participating health plan).

Table B1. Antiarrhythmic Medications of Interest

Antiarrhythmic medications of interest, oral products*
amiodarone hydrochloride†
disopyramide phosphate
dofetilide (as branded Tikosyn)
dronedarone (as branded Multaq)
flecainide acetate
mexiletine hydrochloride
propafenone hydrochloride
propafenone hydrochloride, ER
procainamide
quinidine gluconate, ER
quinidine sulfate
quinidine sulfate, ER
sotalol hydrochloride
sotalol hydrochloride afib/afl§
*Oral products identified by National Drug Codes (NDCs) † While a Healthcare Common Procedure Coding System (HCPCS) code exists for amiodarone (J0282), it is specific to the injection § Generic form of Betapace AF, a sotalol-containing product specifically indicated for use in atrial fibrillation and flutter Abbreviations: ER = extended release; afib = atrial fibrillation; afl = atrial flutter

II. Example Baseline Table Describing the Identified Cohorts of Interest

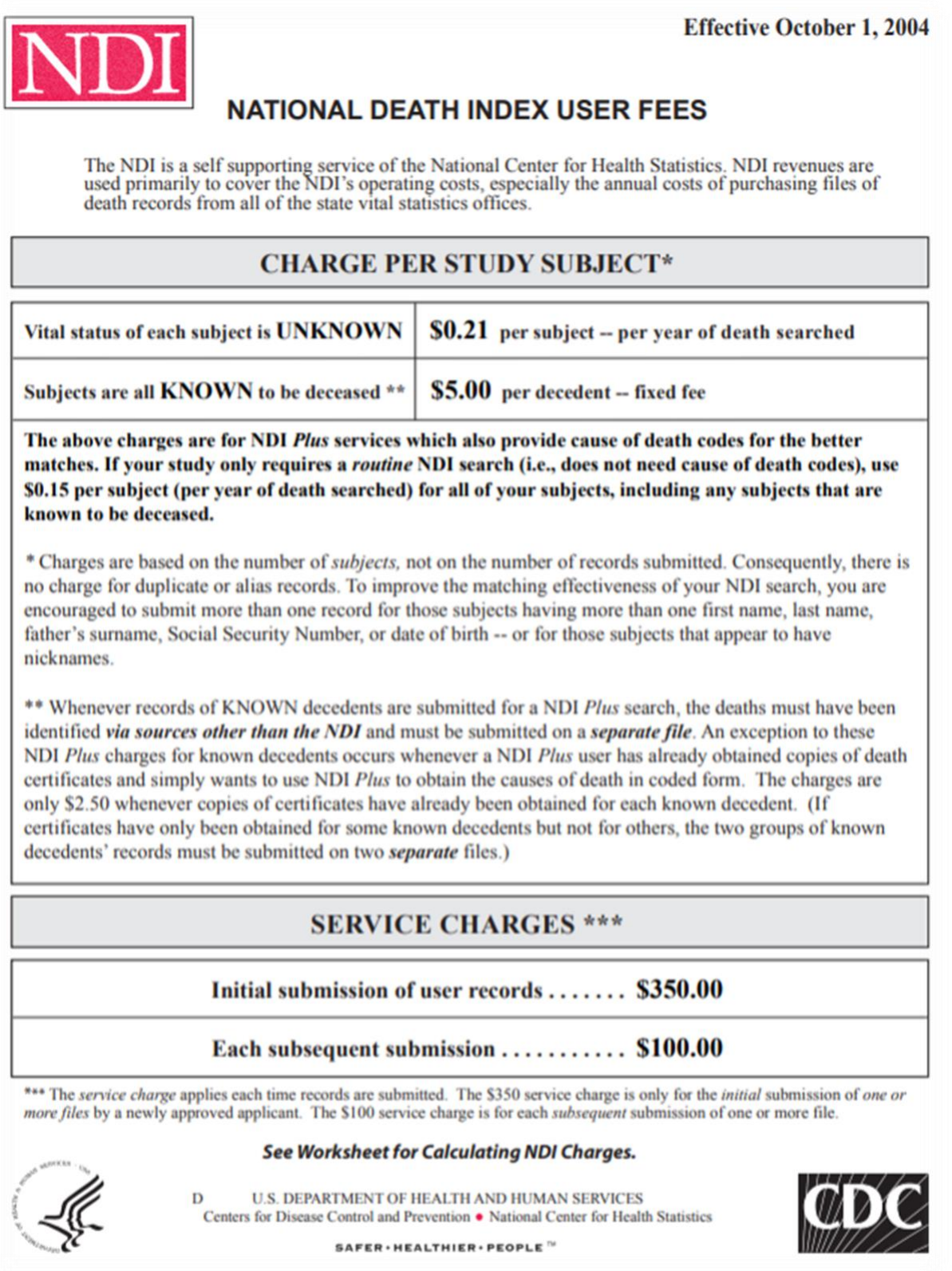
Table B2. Example Baseline Table, Generated by Program 1 During the Pilot Project

Variable	Cohort 1 (Exposed Cohort)	Cohort 2 (Baseline or Comparator Cohort)
Demographics		
Male		
45-54 years	N (% of all males within cohort)	N (% of all males within cohort)
55-64 years	N (% of all males within cohort)	N (% of all males within cohort)
65-74 years	N (% of all males within cohort)	N (% of all males within cohort)
75-85 years	N (% of all males within cohort)	N (% of all males within cohort)
86+ years	N (% of all males within cohort)	N (% of all males within cohort)
Female		
55-64 years	N (% of all females within cohort)	N (% of all females within cohort)
65-74 years	N (% of all females within cohort)	N (% of all females within cohort)
75-85 years	N (% of all females within cohort)	N (% of all females within cohort)
86+ years	N (% of all females within cohort)	N (% of all females within cohort)
Age – Overall		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Age – Male		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Age – Female		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Sex, % female	%	%
Healthcare utilization intensity measures during the baseline period		
Hospitalizations		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Emergency department visits		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Ambulatory care visits		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99

Variable	Cohort 1 (Exposed Cohort)	Cohort 2 (Baseline or Comparator Cohort)
Min	99	99
Max	999	999
Institutional visits		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Other ambulatory care visits		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Unique medications dispensed		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Comorbid conditions, identified during the baseline period		
Arrhythmia / conduction disorder, by type	N (% of cohort)	N (% of cohort)
Atrial fibrillation and flutter	N (% of cohort)	N (% of cohort)
Paroxysmal ventricular tachycardia	N (% of cohort)	N (% of cohort)
Ventricular fibrillation and flutter	N (% of cohort)	N (% of cohort)
Paroxysmal supraventricular tachycardia)	N (% of cohort)	N (% of cohort)
Unspecified paroxysmal tachycardia	N (% of cohort)	N (% of cohort)
Premature beats	N (% of cohort)	N (% of cohort)
Other specified or unspecified cardiac dysrhythmia	N (% of cohort)	N (% of cohort)
Cerebrovascular disease	N (% of cohort)	N (% of cohort)
Coronary heart disease	N (% of cohort)	N (% of cohort)
Diabetes mellitus	N (% of cohort)	N (% of cohort)
Heart failure / cardiomyopathy	N (% of cohort)	N (% of cohort)
Cardioverter-defibrillator/pacemaker	N (% of cohort)	N (% of cohort)
Hyperlipidemia	N (% of cohort)	N (% of cohort)
Hypertension	N (% of cohort)	N (% of cohort)
Kidney disease	N (% of cohort)	N (% of cohort)
Circulatory system disease	N (% of cohort)	N (% of cohort)
Seizure disorder	N (% of cohort)	N (% of cohort)
Smoking	N (% of cohort)	N (% of cohort)
Obesity	N (% of cohort)	N (% of cohort)
Combined Charlson/ Elixhauser co-morbidity score		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	99	99
Co-morbidity score		
Low-0	N (% of cohort)	N (% of cohort)
1	N (% of cohort)	N (% of cohort)
2-3	N (% of cohort)	N (% of cohort)
4-7	N (% of cohort)	N (% of cohort)
8+	N (% of cohort)	N (% of cohort)

Variable	Cohort 1 (Exposed Cohort)	Cohort 2 (Baseline or Comparator Cohort)
Risk of Torsades de pointes (TdP), per Credible Meds		
Known Risk (K)	N (% of cohort)	N/A
Possible risk (P)	N (% of cohort)	N/A
Conditional Risk (C)	N (% of cohort)	N/A
Not classified (N)	N (% of cohort)	N/A
Cohort entry year (Year of Index Date)		
2000	N (% of cohort)	N (% of cohort)
2001	N (% of cohort)	N (% of cohort)
2002	N (% of cohort)	N (% of cohort)
2003	N (% of cohort)	N (% of cohort)
2004	N (% of cohort)	N (% of cohort)
2005	N (% of cohort)	N (% of cohort)
2006	N (% of cohort)	N (% of cohort)
2007	N (% of cohort)	N (% of cohort)
2008	N (% of cohort)	N (% of cohort)
2009	N (% of cohort)	N (% of cohort)
2010	N (% of cohort)	N (% of cohort)
2011	N (% of cohort)	N (% of cohort)
2012	N (% of cohort)	N (% of cohort)
2013	N (% of cohort)	N (% of cohort)
2014	N (% of cohort)	N (% of cohort)
2015	N (% of cohort)	N (% of cohort)
2016	N (% of cohort)	N (% of cohort)
2017	N (% of cohort)	N (% of cohort)

III. NDI Pricing Worksheet



Effective October 1, 2004

NATIONAL DEATH INDEX USER FEES

The NDI is a self supporting service of the National Center for Health Statistics. NDI revenues are used primarily to cover the NDI's operating costs, especially the annual costs of purchasing files of death records from all of the state vital statistics offices.

CHARGE PER STUDY SUBJECT*

Vital status of each subject is **UNKNOWN**	**$0.21** per subject -- per year of death searched
Subjects are all **KNOWN** to be deceased **	**$5.00** per decedent -- fixed fee

The above charges are for NDI *Plus* services which also provide cause of death codes for the better matches. If your study only requires a *routine* NDI search (i.e., does not need cause of death codes), use $0.15 per subject (per year of death searched) for all of your subjects, including any subjects that are known to be deceased.

* Charges are based on the number of *subjects,* not on the number of records submitted. Consequently, there is no charge for duplicate or alias records. To improve the matching effectiveness of your NDI search, you are encouraged to submit more than one record for those subjects having more than one first name, last name, father's surname, Social Security Number, or date of birth -- or for those subjects that appear to have nicknames.

** Whenever records of KNOWN decedents are submitted for a NDI *Plus* search, the deaths must have been identified ***via sources other than the NDI*** and must be submitted on a ***separate file***. An exception to these NDI *Plus* charges for known decedents occurs whenever a NDI *Plus* user has already obtained copies of death certificates and simply wants to use NDI *Plus* to obtain the causes of death in coded form. The charges are only $2.50 whenever copies of certificates have already been obtained for each known decedent. (If certificates have only been obtained for some known decedents but not for others, the two groups of known decedents' records must be submitted on two ***separate*** files.)

SERVICE CHARGES ***

Initial submission of user records $350.00
Each subsequent submission $100.00

*** The *service charge* applies each time records are submitted. The $350 service charge is only for the *initial* submission of *one or more files* by a newly approved applicant. The $100 service charge is for each *subsequent* submission of one or more file.

See Worksheet for Calculating NDI Charges.

D U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention • National Center for Health Statistics

SAFER • HEALTHIER • PEOPLE™

Instructions for calculating NDI charges *(see worksheets on next 2 pages)*

1. If you are submitting more than one file, you may record your calculations for each file on one worksheet or use a *separate* worksheet for each file. Do not include duplicate or alias records in your fee calculations. Just add a footnote to the "Comments" section (in the lower left corner of the worksheet) stating the number of duplicate or alias records associated with each file. A *sample* worksheet appears on the next page, followed by a blank worksheet for your own hand written calculations. Instructions for submitting your records appear on the last page.

2. Subjects with UNKNOWN vital status:

NDI *Plus* (provides cause of death codes) . . . . . $0.21 per subject per year searched
NDI Routine (no cause of death codes). . . . . $0.15 per subject per year searched

a) When all subjects need to be searched against the ***same* range of years**, you only need one row to calculate the charges.

b) When different subgroups of study subjects need to be searched against ***different* ranges of years** (which assumes that different subgroups of subjects were last known to be alive in different years):

Use the first row of the worksheet for that subgroup which needs to be searched for the ***greatest*** number of years, then work toward the present by using the subsequent rows (for records which need to be searched for fewer and fewer years; e.g., 1979-2002, 1980-2002, 1981-2002, etc.).

When creating your file of study subjects, you are encouraged (in most instances) to submit all your subjects' records on one file and to use the Optional User Data field (positions 92-97) to record the date or year last known alive. Even though the NDI will actually search your entire file of records against the greatest range of appropriate years, you will only be charged for the years which needed to be searched for each subgroup -- based on your worksheet calculations.

3. Subjects who are KNOWN decedents:

NDI *Plus* (provides cause of death codes):

$5.00 per decedent -- when researcher has *no* death certificates
$2.50 per decedent -- when researcher has obtained death certificates

NDI Routine (no cause of death codes) . . . . . $0.15 per decedent

Use the last row of the worksheet to calculate charges for **KNOWN** decedents.

For NDI *Plus* searches, known decedents should be submitted on a separate file. The charge for each subject will be a flat $5.00, regardless of the number of years all the records need to be searched against.

For routine NDI searches, records of known decedents can be included in a file of persons with unknown vital status. The charge for each known decedent will be a flat $0.15, regardless of the number of years all the records need to be searched against.

When creating your file of study subjects, you are encouraged to use the Optional User Data field (positions 92-97) to record the date or year of death for each known decedent.

WORKSHEET FOR CALCULATING NATIONAL DEATH INDEX CHARGES

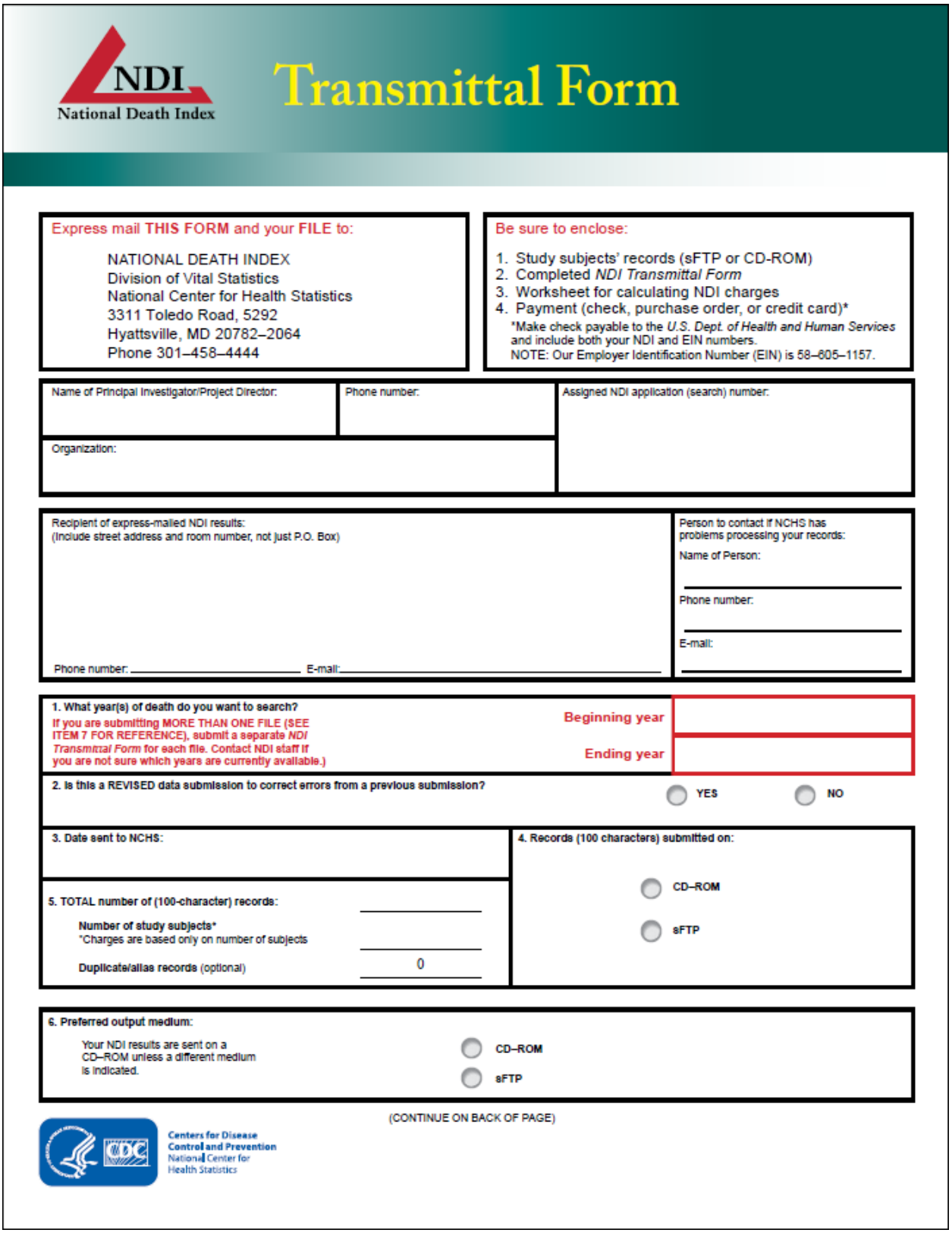
Refer first to instructions and user fees. This worksheet is useful when submitting ***different*** subgroups of records ***on one file*** but only paying for the relevant years each subgroup needs to be searched.

Years Searched	Number of Years		Number of Subjects		NDI Fee (see above)		NDI Charges (for each subgroup)
1992-2002	11	x	100	x	.21	=	231.00
1993-2002	10	x	57	x	.21	=	119.00
1994-2002	9	x	80	x	.21	=	151.20
1995-2002	8	x	110	x	.21	=	184.80
1996-2002	7	x	65	x	.21	=	95.55
1997-2002	6	x	41	x	.21	=	51.66
1998-2002	5	x	72	x	.21	=	75.60
1999-2002	4	x	38	x	.21	=	31.92
2000-2002	3	x	27	x	.21	=	17.01
2001-2002	2	x	22	x	.21	=	36.96
2002	1	x	10	x	.21	=	2.10
		x		x		=	
		x		x	Subtotal	=	997.50
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
		x		x		=	
KNOWN DECEDENTS	1 year	x	200	x	5.00	=	1,000.00

Comments: Duplicate Records
UnKnown = 45
Known = 73

Total record charges	1,997.50
Service charge	350.00
Total NDI charges	2,347.50

IV. NDI Transmittal Form



NDI
National Death Index

Transmittal Form

Express mail **THIS FORM** and your **FILE** to:

NATIONAL DEATH INDEX
Division of Vital Statistics
National Center for Health Statistics
3311 Toledo Road, 5292
Hyattsville, MD 20782–2064
Phone 301–458–4444

Be sure to enclose:

1. Study subjects' records (sFTP or CD-ROM)
2. Completed *NDI Transmittal Form*
3. Worksheet for calculating NDI charges
4. Payment (check, purchase order, or credit card)*

*Make check payable to the *U.S. Dept. of Health and Human Services* and include both your NDI and EIN numbers.
NOTE: Our Employer Identification Number (EIN) is 58–605–1157.

Name of Principal Investigator/Project Director:

Phone number:

Assigned NDI application (search) number:

Organization:

Recipient of express-mailed NDI results:
(Include street address and room number, not just P.O. Box)

Phone number: _____ E-mail: _____

Person to contact if NCHS has problems processing your records:
Name of Person: _____
Phone number: _____
E-mail: _____

1. What year(s) of death do you want to search?
If you are submitting MORE THAN ONE FILE (SEE ITEM 7 FOR REFERENCE), submit a separate *NDI Transmittal Form* for each file. Contact NDI staff if you are not sure which years are currently available.)

Beginning year

Ending year

2. Is this a REVISED data submission to correct errors from a previous submission?

○ YES ○ NO

3. Date sent to NCHS:

4. Records (100 characters) submitted on:

○ CD–ROM
○ sFTP

5. TOTAL number of (100-character) records: _____

Number of study subjects* _____
*Charges are based only on number of subjects

Duplicate/alias records (optional) 0

6. Preferred output medium:

Your NDI results are sent on a CD–ROM unless a different medium is indicated.

○ CD–ROM
○ sFTP

(CONTINUE ON BACK OF PAGE)

Centers for Disease Control and Prevention
National Center for Health Statistics

SUBMITTING YOUR RECORDS AND PAYING FOR NDI SERVICES

Preparing your study subjects' records

- Your records must be put in an ASCII text file format.
- Each record must be 100 positions in length.
- Before sending your file to NCHS, please confirm that each data item (e.g., first name, last name, Social Security Number, date of birth, etc.) begins in its proper position.

Payment for NDI services

You are responsible for accurately calculating your NDI charges. Call the NDI staff on 301-458-4444 if you would like us to confirm your calculations. The following payment options are available:

- **Check** or **purchase order**: Make your check or purchase order payable to -- ***U.S. Department of Health and Human Services.***
- **Interagency agreement**: If you are charging your NDI services to an interagency agreement, specify the name of the agency and the project officer on the back of the *NDI Transmittal Form* that accompanies your data submission. (If your organization is not part of the agency that established the agreement, attach a copy of the agency's letter or e-mail authorizing you to charge your NDI services to the agency's agreement.)
- **Letter of credit**: If you received a credit for future NDI services as a result of making an *advanced payment* or making an *overpayment* for a previous NDI search, use the back of the *NDI Transmittal Form* to indicate that you are applying that credit to cover all or some of your current NDI services. Attach the NDI letter you received which confirms the amount of your credit.

Submitting your records

- Express mail your records to:

  NATIONAL DEATH INDEX
  Attention: Michelle Goodier
  National Center for Health Statistics
  3311 Toledo Road, Room 7318
  Hyattsville, Maryland 20782
  Phone: 301-458-4444

- Your submission should include:

  1. Diskette or CD-ROM (containing your subjects' records)
  2. *NDI Transmittal Form*
  3. *Worksheet for Calculating NDI Charges*
  4. Check or purchase order (or payment instructions as specified above)

04-0385 (6/04)

V. Example Memo describing rationale for utilizing Social Security Numbers when linking databases to the National Death Index

Memo, the National Death Index (NDI) and required matching variables

The National Death Index (NDI) is currently the only complete national source of death and cause of death information accessible to population-based epidemiologic studies in the United States. The NDI, a self-supporting service within the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention, is a centralized database of death record information compiled from state vital statistics offices. The NDI provides death information including death date as well as cause of death information (NDI+ data).[1]

There are a number of variables that are required in order to match a requester's record with a death record within the NDI. These include names, dates of birth (or components of the date such as only year and month), marital status, father's surname (for women), and Social Security Number (SSN). The SSN is a key matching variable required by the NDI for matching to a database of death records so that death information can be obtained.[1] The NDI requires any record submitted for matching to contain at least one of the following combinations of data elements before a search will be conducted[2]:

- First AND last name AND **Social Security Number**
- **Social Security Number**, AND month AND day AND year of birth, AND sex
- First AND last name AND month AND year of birth

NDI record matches are ranked based on the number of data items, as well as parts of data items (e.g., name components) that agree, with a record containing an agreement on SSN always being ranked first. [1] If a SSN is not submitted, NDI classifies the match as a lower quality class automatically, indicating lower confidence in the quality of the match.[2] Studies have indicated SSN number is a necessary, but not sufficient, requirement for a match (i.e., SSN and other information such as sex, date of birth, name are also required for the NDI matching algorithm and producing high quality matches).[3,4]

While it is possible to submit name and date of birth without SSN to the NDI for matching, any matches are automatically considered to be of lower quality, which can lead to misclassification. Misclassification in this instance means that patients could be misclassified as 'dead' when they are alive, or the NDI could be unable to locate a death record for a patient who is actually dead, thus incorrectly indicating that a patient may still be alive. Additionally, the NDI may return multiple death records that match a requester's submitted record. When multiple matching potential death records are returned, it is critical to evaluate which of the multiple matches is the "true" matching record, a process that requires complicated algorithms. Given the risk of misclassification or selecting incorrect matches, evaluating matches which are flagged as low-quality is recommended, but requires more resource intensive manual review steps.

For these reasons, submission of SSN, when it is available, along with other key NDI required information is important for accurately and efficiently linking patients in large population-based epidemiologic studies to data from the NDI. This approach is consistent with many other NDI linkage studies.[4-6]

References

1. National Center for Health Statistics Fact Sheet: National Death Index, August 2017 2017; https://www.cdc.gov/nchs/data/factsheets/factsheet_ndi.htm. Accessed March 28, 2018, 2018.
2. Centers for Disease Control and Prevention. National Death Index User's Guide. https://www.cdc.gov/nchs/data/ndi/ndi_users_guide.pdf. Accessed April 13, 2018.
3. Fillenbaum GG, Burchett BM, Blazer DG. Identifying a national death index match. *Am J Epidemiol.* 2009;170(4):515-518.
4. Wojcik NC, Huebner WW, Jorgensen G. Strategies for using the National Death Index and the Social Security Administration for death ascertainment in large occupational cohort mortality studies. *Am J Epidemiol.* 2010;172(4):469-477.
5. Skopp NA, Smolenski DJ, Schwesinger DA, Johnson CJ, Metzger-Abamukong MJ, Reger MA. Evaluation of a methodology to validate National Death Index retrieval results among a cohort of U.S. service members. *Ann Epidemiol.* 2017;27(6):397-400.
6. Curtis MD, Griffith SD, Tucker M, et al. Development and Validation of a High-Quality Composite Real-World Mortality Endpoint. *Health Serv Res.* 2018.

VI. NDI Death and Cause of Death Tables

A. Structure of the NDI Death Table Saved by Distributed Program 6

Table B3 illustrates the structure of the **NDI Death Table** (i.e., NDI_Death) which is created by Distributed Program 6. The NDI Death Table includes only one record or 'row' per patient. The following variables were taken directly from the Sentinel Common Data Model (SCDM) Death Table:[34] patient ID, death date, imputed death date (not relevant to NDI data, but is included to ensure SCDM model compliance), source (e.g., NDI), and confidence (e.g., excellent). We expanded the NDI Death Table to include state of death code (list of valid codes included in Table B4) and death certificate number as these variables are available within NDI records, but are not currently included in the SCDM. Distributed program 6 populates these variables as outlined in the Table B3 'Comments' column.

Table B3. Structure of the National Death Index (NDI) Death Table saved by Distributed Program 6

Variable	SAS Type (Length)	Format	Label	Valid Values	Included in the Sentinel Common Data Model?	Comments
PatID	Char(#)	Site specific length	Patient ID	Alpha-Numeric	Yes	Arbitrary person-level identifier, used to link across tables.
DeathDt	Num(4)	mmddyy10.	NDI Date of death	Valid SAS date	Yes	Death date on the NDI record
DtImpute	Char(1)	$1.	Imputed death date	N = Not imputed	Yes	Defaulted to DtImpute='N' as not relevant to records attained from the NDI
Source	Char(1)	$1.	Source of death information	N = National Death Index	Yes	Defaulted to Source='N'
Confidence	Char(1)	$1.	Confidence	E = Excellent	Yes	Defaulted to Confidence='E'
State_Death_Code	Char(3)	$3.	State of death code	See Table B4	No	State of death code is provided on the NDI record
Death_Cert	Char(6)	$6.	Death certificate number	Free text	No	Jurisdiction death certificate number is provided on the NDI record

Table B4. State Postal Codes

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Alabama	AL	01	New Jersey	NJ	31
Alaska	AK	02	New Mexico	NM	32
Arizona	AZ	03	New York	NY	33
Arkansas	AR	04	New York City	*n/a*	33C
California	CA	05	North Carolina	NC	34
Colorado	CO	06	North Dakota	ND	35
Connecticut	CT	07	Ohio	OH	36
Delaware	DE	08	Oklahoma	OK	37
District of Columbia	DC	09	Oregon	OR	38
Florida	FL	10	Pennsylvania	PA	39
Georgia	GA	11	Rhode Island	RI	40
Hawaii	HI	12	South Carolina	SC	41
Idaho	ID	13	South Dakota	SD	42
Illinois	IL	14	Tennessee	TN	43
Indiana	IN	15	Texas	TX	44
Iowa	IA	16	Utah	UT	45
Kansas	KS	17	Vermont	VT	46
Kentucky	KY	18	Virginia	VA	47
Louisiana	LA	19	Washington	WA	48
Maine	ME	20	West Virginia	WV	49
Maryland	MD	21	Wisconsin	WI	50
Massachusetts	MA	22	Wyoming	WY	51
Michigan	MI	23	Puerto Rico	PR	52
Minnesota	MN	24	Virgin Islands	VI	53
Mississippi	MS	25	Guam	GU	54
Missouri	MO	26	Canada	CN	55
Montana	MT	27	Cuba	CU	56
Nebraska	NE	28	Mexico	MX	57
Nevada	NV	29	Remainder of World	RW	59
New Hampshire	NH	30	Unknown		99

B. Structure of the National Death Index (NDI) Cause of Death (COD) Table Saved by Distributed Program 6

Table B5 illustrates the structure of the **NDI COD Death Table** (i.e., NDI_COD), which also leverages the SCDM and includes one row per COD from the NDI data. The following variables were taken directly from the SCDM COD Table: Patient ID, source, confidence, cause of death code, code type (e.g., ICD-9, ICD-10), cause type (e.g., underlying), source (e.g., state death files, and confidence (e.g., excellent). However, we expanded the accepted values for the SCDM variable 'cause type' variable to accommodate NDI data and added several NDI provided variables within the Table. Many of these NDI variables are highly specific to the National Center of Health Statistics (NCHS), and the NDI and NCHS provide some basic information about the variable structure within the *NDI User's Guide*[18] with additional details included on the Centers for Disease Control and Prevention (CDC) vital statistics website.[41–43] As it is possible future studies might be interested in utilizing these additional NDI variables, we prioritized saving all returned NDI data variables and cited these NDI and CDC resources in this document so that future studies can examine these resources as needed. Thus we also included several additional NDI variables in the **NDI COD table**: variables indicating COD physical position on a death certificate[44] (e.g., Part and Line number), sequence of CODs on a death certificate,[44] and several flags indicating reordering of CODs or reassignment of an underlying COD based on NCHS programs used to process death certificate data.[41–43]

NCHS programs used to process death certificate data are intended to standardize death certificate coding.[41–43] These programs provide a number of functions including: 1) automating cause coding rules and assigning ICD codes to each COD text entity on a death certificate, 2) applying World Health Organization (WHO) rules to these assigned ICD codes and selecting an underlying COD in a standardized fashion, 3) processing CODs in manner which best describes the overall medical certification portion of the death certificate. The NDI returns CODs to end users with flags for an 'entity COD', a 'record axis COD', or an 'underlying COD', which allow end users to investigate COD coding as needed.

Entity COD flags are the original positions of COD codes on the death certificate and the original conditions coded for selection of the underlying cause of death.[18,41] Record axis COD flags are created from the original set of entity CODs and NCHS considers these CODs as those that best describe the overall medical certification portion of the death certificate.[41,43] All of these variables are returned to end users for analyses and can be used to identify CODs on Part I and Part II of the death certificate.[44] Distributed SAS program 6 populates variables included in the NDI Death COD Table from the NDI data as outlined in the **Table B5** 'Comments' column. Please note that many of these are NDI specific variables, and while we extracted some relevant information about entity and record axis CODs below, please refer to NDI and NCHS documentation for additional details.[18,41–43]

Entity CODs: Entity CODs provide the coded COD and the location on the death certificate[44] in the exact manner given by the original medical certifier and are the original conditions coded for selection of the underlying cause of death. However, the NCHS states this approach may be limited and describes, *"The original scheme for coding conditions contained on the death certificate was designed with two objectives in mind. First, to facilitate etiological studies of the relationships among conditions, it was necessary to reflect accurately in coded form each condition and its location on the certification in the exact manner given by the certifier. Secondly, the codification needed to be carried out in a manner by which the underlying cause of death could be assigned through computer applications."* [42] This general approach is hereafter called entity coding.

Record axis COD: Record axis COD (i.e., record axis codes) are created from the original set of entity CODs, which the NCHS refers to as 'record axis multiple cause data'. NCHS states, "*Unfortunately, the set of multiple cause codes produced by entity coding is not conducive to a third objective—the generation of person-based multiple cause statistics. Person-based analysis requires that each condition be coded within the context of every other condition on the same certificate and modified or linked to such conditions as provided by ICD–9. By definition, the entity data cannot meet this requirement because the linkage provisions distort the character and placement of the information originally recorded by the certifier. Essentially, the axis of classification has been converted from an entity basis to a record (or person) basis. The record axis codes are assigned in terms of the set of codes that best describe the overall medical certification portion of the death certificate. The translation is accomplished by a computer system called TRANSAX (TRANSLATION OF AXIS) through selective use of traditional linkage and modification rules for mortality coding. Underlying cause linkages which simply prefer one code over another for purposes of underlying cause selection are not included. Each entity code on the record is examined and modified or deleted as necessary to create a set of codes which are free of contradictions and are the most precise within the constraints of ICD-9 and medical information on the record. Repetitive codes are deleted. The process may (1) combine two entity axis categories together to a new category thereby eliminating a contradiction or standardizing the data; or (2) eliminate one category in favor of another to promote specificity of the data or resolve contradictions.*" [42] Although in this language, NCHS references ICD-9 coded causes of death, this explanation also applies to ICD-10 coded causes of death.

*Table B5. Structure of the National Death Index (NDI) Cause of Death (COD) Table saved by Distributed Program 6**

Variable	SAS Type (Length)	SAS Format	Label	Valid Values	Included in the Sentinel Common Data Model?	Comments
PatID	Char(#)	$#.	Patient ID	Alpha-Numeric	Yes	#= Site specific length
Source	Char(1)	$1.	Source	N = National Death Index	Yes	Defaulted to Source='N'
Confidence	Char(1)	$2.	Confidence	E=Excellent	Yes	Defaulted to Confidence='E'
COD	Char(8)	$8.	Cause of death	Free text	Yes	Any COD code included on a death certificate
CodeType	Char(2)	$2.	Code type	09 = ICD-9 10 = ICD-10	Yes	This model can be expanded as needed (e.g., ICD-11)
CauseType ±	Char(3)	$3.	Cause type	U = Underlying C = Contributory I = Immediate/Primary O= Other R=Record Axis COD 358 = Indicates recoding of original cause code 113 = Indicates recoding of original cause code 130 = Indicates recoding of original cause code	Yes, but we expanded this table to include the values 'O', 'R', '358', '113', and '130' are values which allow data returned by the NDI to be saved.	• CauseType='U' is provided directly by the NDI • CauseType="C" for any Entity Part II listed cause on a death certificate • CauseType= 'I' indicates the Entity cause of death written on the first line in Part I of a death certificate • CauseType="O" for any other Entity Part I listed cause on a death certificate • CauseType='R' indicates Record Axis CODs, which is based on a possible reclassification using ACME and TRANSAX software. Please see NDI documentation for details[2] • CauseType= '358', '113', or '130' indicate a recoding of the original ICD-9 or ICD-10 death certificate cause code into new groups for use in NCHS publications
CauseCat	Char(1)	$1.	Cause category	E=Entity COD R=Record axis COD U=Any underlying*	No	• NDI returns CODs as either an entity COD 'E', a record COD 'R', or an underlying COD 'U' • Entity CODs are based on the physical COD position on a death certificate and one of these is an underlying COD • Record CODs are based on classification of CODs using National Centers for Health Statistics (NCHS) Mortality Medical Data System (MMDS) and ACME and TRANSAX software.[3,5] Please see NDI documentation[2] and NCHS software documentation for details[2,5]

Variable	SAS Type (Length)	SAS Format	Label	Valid Values	Included in the Sentinel Common Data Model?	Comments
Position1[±]	Char(1)	$1.	Cause position	Part/line number on death certificate 1=Part I, line 1 (a) 2=Part I, line 2 (b) 3=Part I, line 3 (c) 4=Part I, line 4 (d) 5=Part I, line 5 (e) 6=Part II R=Record cause category U=Any underlying	No	• Position allows for reconstruction of the COD position and line number on a death certificate • Position= 1-6 is coded when CauseCat='E' • Position='R' when CauseCat='R'[*] • Position='U' when the CauseType='U', or when recoding has occurred (e.g., CauseType='358', '113', or '130')[†]
Position2[±]	Num(3)	1.	Position Sequence	1-7 or missing	No	• Position allows for reconstruction the COD position and line number on a death certificate • Position 2 is the sequence of condition within part/line code, and values are taken directly from NDI data.
Cause_Sequence	Num(3)	2.	Sequence Number	1-20 or missing	No	• Up to 20 cause of death codes are available for each of entity COD and record COD, this variable provides the overall sequence, please see the NDI users guide for more details[2]
Nature_Injury	Char(1)	$1.	Nature of Injury Code	1= ICD–9 nature of Injury code 0=All other codes N=Not applicable	No	• The *NDI User's Guide* lists this variable as being returned with all NDI records, a value of 1 is permitted only when CodeType=09 and a relevant nature of injury code is provided. • N is for Entity codes only

[*] NDI returns CODs as either an entity, a record COD, or an underlying COD. Entity CODs are based on the physical position on a death certificate and one of these is an underlying COD. Record CODs are based on a possible reclassification using ACME and TRANSAX software[3] and there is also assignment as underlying CODs. As the Cause Type variable can have multiple values to indicate an underlying COD (e.g., the 358, 113, and 130 types), variables 'CauseType' and 'CauseCat' enable full classification of each NDI COD variables.

[†] A recode of the ICD–10 underlying COD into one of 358, 113, or 130 groups for use in NCHS publications. NDI records for deaths in 1998 or earlier will have CODs using ICD-9 and the 3 groups are 282, 72, and 61. Specifically, the NDI states that '61' or '131' groups indicate an 'Infant Cause Recode' for use in NCHS publications, and that this is only applicable when compiling statistics on infants under 1 year of age. Please see NDI documentation for details.[2]

± Please see NDI User's Guide[2], page 34 item 4 to learn more about position codes. The Guide states, 'space has been provided for a maximum of 20 conditions. Each condition takes 7 positions in the record. Records that do not have 20 conditions are blank in the unused positions or area.

VII. Lessons Learned from Piloting a Distributed Technical Workflow

This section summarizes technical challenges and recommendations to assist future studies with efficiently achieving linkage to data from the NDI.

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
1	NDI File Submission (Program 4)	Format of 2018 files for Early Release NDI transmission	NDI stated that early release requests should be broken out and submitted as a separate file, rather than just one file containing all patients included in the initial data request (i.e., only provide the NDI records that should be searched within early release files). NDI offered to make the change on their end as a favor, but future studies should be aware of this request.	Future projects should separate patient files to match with routine searches vs. patient files to match with early release file(s).	1
2	NDI File Return (After Program 4, prior to implementing Program 5)	Data file names	We noted a discrepancy between file names described in the *NDI User's Guide* (page 14), and names of files actually returned to Study Sites. The *NDI User's Guide* (page 14) lists 5 files that get returned in two groups:	For multi-site distributed linkage studies where computer programs expect specific file names, understanding and updating programs to reflect the NDI's actual file names is critical. For a single site project, the filenames can be managed when they are returned.	1

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
			1) 'Match', 'NoMatch', and 'Rejects': These contain exact records as submitted, with the NDI matching result. 2) 'Combined' and 'Cause': For matching records only, these files contain matching variables as well as, in the Cause file, causes of death. The actual file names followed this pattern for routine searches: YYYY-A999A99.MATCH, YYYY-A999A99.NOMATCH, YYYY-A999A99.REJECTS, 2018-X010F00.COMBINED, and 2018-X010F00.CAUSE, where: • YYYY= the latest year searched • A999A99= the project number assigned by the NDI Note that there are no file extensions.	Note that some of this specificity wasn't included in the *NDI User's Guide*, and work closely with the NDI as needed.	

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
			For files returned based on searches of NDI Early Release files, a prefix of "ER" was placed on each file name; for example: ER2018-X010F00.CAUSE.		
3	Inception (All Programs)	Programming for matching and analysis	It is most efficient to develop a full NDI linkage plan tailored to each study in advance, as it avoids need for re-work. At project initiation all programming packages should be pre-specified where possible and updated as plans change throughout the study, and analytic plans should be tailored for each individual study. This will avoid rework.	Develop a full analytical plan followed by programming steps to ensure: 1. Each programming package includes required tables and variables. 2. Tables utilized from package to package are specified.	1

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
4	Inception (All Programs)	Patient attrition	During the pilot project, we learned at Study Sites some patients transferred into Health Plans which were deemed 'no-longer eligible for research', meaning that there was patient attrition over time. To track patient attrition throughout the study, we retrofit some packages in places that were not optimal.	Future studies should ensure all program packages track and manage patient attrition across time and create plans for how to manage this attrition in their study analytic plans.	1
5	NDI File Submission (After Program 4)	Version of WinSCP application if important for file transfer	One Study Site identified older versions of WinSCP prevent connection with the NDI server, hindering the ability to transfer files using the SFTP method. Follow-up from Health Plan: "Unfortunately, we do not know the minimum version requirement for the NDI SFTP to function. The update we received was for Version 5.13.1 (Build 8265)." (version 5.13.1 verified to be acceptable).	Future projects should ensure Study Sites install the latest version of WinSCP (or any other software NDI requires in the future) and connect with NDI to confirm acceptable version.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
6	NDI recommendations for accepting matches (Program 5)	Impact of Using Alternative Cutoff Scores table	This table appears in the *NDI User's Guide* (page 54) and contains statistics on utilizing alternative cutoff scores. Some of the percentages in the table cannot be calculated with the data provided in the guide, which led to some confusion. The NDI let us know that some of these challenges are due to rounding (i.e., original number values are not provided in the *NDI User's Guide*)	Interpret this table with care, and work closely with the NDI as required.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
7	Recommendations for defining 'ties' between NDI records (Program 5)	Definition of 'ties' when selecting the 'best' NDI match	In our pilot study, we defined a tie between potential NDI matches as two or more records with an exact probabilistic matching score and ultimately did not identify any 'ties'. However, subsequently we learned that some studies also consider relaxing the definition of 'ties' to include potential matches within a range of the probabilistic matching scores (rather than requiring an exact match).	Future studies could consider using a less stringent approach for defining ties and conduct manual review to identify final matches. This approach may identify more 'tied' NDI records which would require additional manual evaluation by the Study Sites, but also provides opportunities for confirming match selection in situations where multiple potential matches have very similar probabilistic matching scores. In future studies, program 5 could be adjusted to include a broader definition of 'tied' NDI records, as needed. However, please keep in mind reviewing records manually and enforcing uniformity of manual review in a multi-site study can be labor intensive.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
8	Inception (All Programs)	Site death data check	If Study Sites have access to some death data within their source database, future studies should create a plan for if/when/how they plan to use these data. Planning for this in advance will ensure that programs are updated appropriately so that these alternative sources of death data may be used to examine concordance between returned NDI data and deaths already known to the Study Site, thereby assisting with determining methods for selecting the best match.	Plan with sites that contain death data as to: 1. Stage of project they'll be asked to check death data 2. Nature of checking/comparisons 3. Results to be provided to Coordinating Center	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
9	Inception (All Programs)	Site death data check	If Study Sites have access to death data within their source database, it is important to understand the strengths and limitations of these data sources. During the pilot study, some Study Sites examined their state-based record linkage and found deaths might be missed if a patient moves and dies out of state. Otherwise agreement was very good. This context helped with interpretation of NDI linkage results (i.e., deaths NDI identified that were not identified at a Study Site may be due to patients with deaths outside a particular state).	It is important to understand systematic/methodologic differences between NDI and alternative data sources before using it to inform the best match.	2
10	NDI File Submission (After Program 4)	*NDI User's Guide*: File transfer methods	The *NDI User's Guide* does not describe the file transfer methods the NDI will accept (SFTP Methods) and determining current acceptable methods will save Study Sites time, especially in multi-site studies.	We recommend testing SFTP site authentication and file exchange methods ahead of time, and working closely with the NDI as needed.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
11	Assessing NDI matches (Program 5)	*NDI User's Guide*: Understanding NDI designation of a "True" match	The NDI rules for assigning records as 'True' match (i.e., Status_Code 1 or 0) were not entirely clear to the study team after reviewing the *NDI User's Guide,* which led to questions as the Study Team worked to determine how best to select the 'best' NDI match from all returned NDI matches.	Study teams should be prepared to work collaboratively to determined how best to select the 'best' NDI match from all returned NDI matches. If questions come up, consult with the NDI as needed to understand current practices and procedures.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
12	NDI COD Recodes (Program 5)	*NDI User's Guide*: Understanding NDI recodes of cause of death	Studies should be aware that sometimes cause of death is recoded from the original death certificate by the NDI, especially if examining concordance between cause of death codes at the Study Site and NDI cause of death codes. There is some information about this in the *NDI User's Guide*, but the Study Team located other sources of information on the CDC website, and through the NDI project staff. There is recode of the ICD–9 or ICD–10 cause codes into: - 282 or 358 groups - 72 or 113 groups - 61 or 130 infant groups	For examining concordance between cause of death codes at the Study Site and NDI cause of death codes, review the *NDI User's Guide* and work with NDI project staff to specify the groups and how recording is performed.	2

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
13	Throughout (All Programs)	Communications	The importance of maintaining frequent Study Site and Coordinating Center Communications cannot be overemphasized; sometimes the NDI provides unexpected files or information which does change over time, and direct communications between all parties can save a great deal of time and avoid confusion.	Maintain frequent communications between the Coordinating Center and sites to enable sites to have complete information about processes, requirements, timelines, tasks, etc.	2
14	NDI File Return (After Program 4, prior to implementing Program 5)	CD titled “Names Count!” (reference in Chapter 4 of the NDI User’s Guide - pg. 47)	"NDI Names Count!" is no longer available for request per NDI on 29 Jul 2019. The list is utilized to help develop the process for acceptance criteria for matches involving common names.	Future projects should note this file is no longer available..	3

#	**Project Stage:** Stage where lesson appeared	**Lesson topic:** Short name for topic	**Description:** Full description of the lesson	**Recommended Action:** How to ensure the project work moves forward more smoothly	**Relative importance:** 1- High (red), 2- Medium (yellow), 3- Low (green)
15	NDI File Return (After Program 4, prior to implementing Program 5)	Unexpected file, returned to Study Sites by the NDI, ER2018-X010F00.ERSTATUS file	NDI provided one file not described in the *NDI User's Guide*, and with no data dictionary. This file " ER2018-X010F00.ERSTATUS" contains the percentage complete by State of the 2018 NDI early release file. Routine = Demographics and NDI Plus = Cause of Death. NDI confirmed that there is no data dictionary for this file and no information is included in the NDI User Guide.	Future projects should be aware that this file may also be returned.	3
16	NDI entity CODs (Program 5-6)	*NDI User Guide*: Clarity on definition of "entity" conditions	It took research to understand the meaning on entity conditions, and how to translate them for use in our study, as limited information was included in the *NDI User's Guide*.	We learned that that "entity" conditions are based on physical location of CODs on death records, and recommend that future studies consult with the NDI staff and refer to the NCHS website to attain up to date information as needed.	3

DEPARTMENT OF POPULATION MEDICINE
HARVARD MEDICAL SCHOOL | Harvard Pilgrim Health Care Institute | TIDE THERAPEUTICS RESEARCH AND INFECTIOUS DISEASE EPIDEMIOLOGY

Distributed Processes for Attaining Death and Cause of Death Information from the National Death Index (NDI) in Multi-Center Studies

Supplemental Appendix: Distributed Program Specifications Implemented in a Pilot Study

This supplemental appendix includes technical programming specifications implemented in a project which piloted a distributed process for attaining death and cause of death information from the NDI in multi-center studies. We designed the overall distributed NDI linkage process to be flexible and reusable, and provided the general methodology developed and implemented in the pilot project. These specifications were utilized during the pilot study and can be leveraged in future studies that require a distributed NDI linkage process. Technical lessons learned during the pilot study and recommended adjustments to the process are captured in the main report. Please note, modifications may be necessary based on individual study needs.

Contents

Program Package 1-2: Identify the Cohorts of Interest, and Select which Patients to Submit to the NDI for Matching

Overview

This project piloted a distributed process for attaining death and cause of death information in multi-center studies through a use case designed to examine the associations between select antiarrhythmic medications and all-cause mortality and specific causes of death. The project was led and coordinated by the Harvard Pilgrim Health Care Institute (HPHCI), which worked closely with the FDA and participating health plans. Six health plans participated in this project. Although the project leveraged Sentinel infrastructure and was built on the successful collaboration among participating health plans in Sentinel, it was conducted outside of Sentinel.

The first step to conducting an efficient distributed linkage of multiple Study Site databases to NDI data is identifying cohorts of interest at each of the Study Sites. During the pilot study, we identified use case cohorts using the publicly available Sentinel Cohort Identification and Descriptive Analysis (CIDA) program, version 7.0.0 in combination with additional code to provide descriptive information to the study team. The CIDA program is available on the Sentinel website (https://www.sentinelinitiative.org/), and was used to identify the cohorts of interest in our pilot project (i.e., Cohort Identification Package, Program Package 1). In our pilot project, we included two cohorts- antiarrhythmic drug users cohort (i.e., Cohort 1, exposed cohort) and a matched non-users cohort (i.e., Cohort 2, unexposed cohort). We submitted patients with deaths recorded at the study sites, as well as those with potential deaths to the NDI for linkage. We defined potential death as health plan disenrollment between cohort entry and cohort exit plus 365 days, without subsequent re-enrollment or medical or drug utilization >60 days post-disenrollment, but this could be modified by future studies. For more details regarding the use case, please refer to the main report document.

Each study will need to determine exactly which patients to submit to the NDI for linkage and may also consider whether or not to submit all study patients to the NDI for linkage or just study patients whose vital status is unknown. While in some instances it might be preferable to submit all patients captured within a particular study to the NDI for linkage, it is possible that the number of patients eligible for the NDI data linkage may be larger than study resources allow. To address this potential, we built options into the developed NDI linkage process to allow future studies to manage the overall number of records being submitted to the NDI for matching.

The purpose of Program Package 2 described below, is to create filtering logic that will be applied to the patient level data held by a site (i.e., health plan). From the Cohort Identification Package generated with the CIDA program, an aggregate file will be returned for examination (See Identify Exposed and Baseline Cohorts). This file will be examined by a manual process, in order to determine filtering criteria to be applied to the patient level data in Program Package 2. This program package creates a filter, which is the main outcome of this program package (see - Aggregate Data Returned to Analytic Site).

I. Identify Exposed and Baseline Cohorts

The following is the data dictionary of patient-level records for the exposed and baseline cohorts used in the following Program Packages.

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
PatID	Char(##)	$##.	Patient Identifier	Free text	From CIDA MSTR table; site specific lengths
IndexDt	Num(4)	mmddyy10.	Index date of drug therapy	Valid SAS dates	From CIDA MSTR table
Enr_Start	Num(4)	mmddyy10.	Start of enrollment	Valid SAS dates	From CIDA MSTR table; start of enrollment span where index exposure was identified
Enr_End	Num(4)	mmddyy10.	End of enrollment	Valid SAS dates	From CIDA MSTR table; end of enrollment span where index exposure was identified
Year	Num(3)	4.	Index date year	2000+	Year of index date
DrugName	Char(60)	$60.	Name of index drug	Name of a drug *or* "Multiple"	From Description in Appendix A
RawDisp	Num(3)	3.	# unique medications dispensed	0+	From CIDA MSTR table; # of unique dispensing rows; 0 for baseline cohort
AdjustedDisp	Num(3)	3.	Days supplied	0+	From CIDA MSTR table; # of collapsed dispensings; 0 for baseline cohort
TotRxSup	Num(3)	3.	Days supplied	0+	From CIDA MSTR table; # of days supplied; 0 for baseline cohort
DaysSuppliedWeeks	Char(10)	$10.	Days supplied in weeks	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95-99, 100-104, 105+, or "BASE"	Integer value of TotRxSup / 7
TotRxAmt	Num(3)	3.	Amount supplied	0+	From CIDA MSTR table; # of ; amount supplied; 0 for baseline cohort
EpisodeEndDT	Num(4)	mmddyy10.	Exposure episode end date	Valid SAS dates	From CIDA MSTR table
Age	Num(3)	3.	Age at index date	45+	From CIDA MSTR table, set to integer
Birth_Date	Num(4)	mmddyy10.	Date of birth	Valid SAS dates	From CIDA MSTR table

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
Sex	Char(1)	$1.	Sex of patient	M, F only	From CIDA MSTR table; values of A and U will be combined into O (unknown)
Race	Char(1)	$1.	NDI Race of patient	1=White 2=Black 3=Indian[1] 4=Chinese 5=Japanese 6=Hawaiian 7=Other non-White[2] 8=Filipino 9=Unknown 0=Other Asian or Pacific Islander	Populate with NDI values; mapped From SCDM/CIDA MSTR table, as follows: CIDA Table NDI Value 0 0 1 3 2 9 3 2 4 9 5 1 Not all NDI values can be mapped from CIDA values
Hispanic	Char(1)	$1.	Hispanic indicator	N = No U = Unknown Y = Yes	From CIDA MSTR table
State	Char(3)	$2.	State	US postal state codes	From CIDA MSTR table
DeathDt	Num(4)	mmddyy10.	Death date	Valid SAS dates, or special missing .A, if still alive	From CIDA MSTR table
DeathSrc	Char(1)	$1.	Source of death date	D=Death table E=Encounter table B=Both tables A=Alive/no death evidence	Requires CIDA modification or ad hoc code

[1] Indian (includes American, Alaskan, Canadian, or Mexican Indian; Eskimo; and Aleut)

[2] Includes Cajun and Creole

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DeathOccurredFlag	Char(5)	$5.	Time period when death occurred	A=Alive; no death occurred Y_E= Death occurred from IndexDt through EpisodeEndDt Y_365=Death occurred from EpisodeEndDt to EpisodeEndDt + 365 N=Death occurred within some other time interval	
Months_to_Next_NDI_Year	Num(3)	$2.	# of months to reach next year of NDI data	3 ,6, 9, 12, or .A	Special missing value of .A for alive
ExactNumVisit	Num(3)	3.	All medical encounters during the baseline period	0+	From CIDA MSTR table
NumVisits_AV	Num(3)	3.	# ambulatory care visits during baseline	0+	From CIDA MSTR table
NumVisits_OA	Num(3)	3.	# other ambulatory care visits during baseline	0+	From CIDA MSTR table
NumVisits_IP	Num(3)	3.	# of hospitalizations	0+	From CIDA MSTR table
NumVisits_IS	Num(3)	3.	# institutional care visits during baseline	0+	From CIDA MSTR table
NumVisits_ED	Num(3)	3.	# of ED visits during baseline	0+	From CIDA MSTR table
Combined_Score	Char(5)	$5.	Comorbidity score category during baseline	Low-0 1 2-3 4-7 8+	From CIDA MSTR table
Combined_Score_Num	Num(3)	2.	Charleson Elixhauser combined comorbidity score	-2 – 26	From CIDA MSTR table
CCIELIXGRP	Char(10)	$10.	Pre-index comorbidity score category	Low-0 1 2-3 4-7 8+	From CIDA MSTR table

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
NumGeneric	Num(3)	3.	Number of generic names dispensed during the episode	1+ for exposed cohort; 0 for baseline cohort	From CIDA MSTR table
NUMCLASS	Num(3)	3.	Number of unique class names dispensed	1+ for exposed cohort; 0 for baseline cohort	From CIDA MSTR table
Num_Arrhythmia_Any	Num(3)	3.	Arrhythmia / conduction disorder, by type	1= Occurred in baseline period 0= Did not occur	Composite outcome collapsing *any 1/0* from below: distinct counts of: Num_Afib_flutter, Num_PVT, Num_VF_flutter, Num_PST, Num_Unsp_PT, Num_prem_beat, Num_other_dysrhythmia
Num_Afib_Flutter	Num(3)	3.	Atrial fibrillation and flutter	1= Occurred in baseline 0= Did not occur	Renamed from Covar1 (AFIB)
Num_Defrib	Num(3)	3.	Cardioverter-defibrillator/pacemaker	1= Occurred in baseline 0= Did not occur	Renamed from Covar2 (CARPM)
Num_Cerebro	Num(3)	3.	Cerebrovascular disease	1= Occurred in baseline 0= Did not occur	Renamed from Covar3 (CBVD)
Num_CHD	Num(3)	3.	Coronary heart disease	1= Occurred in baseline 0= Did not occur	Renamed from Covar4 (CHD)
Num_Circulatory	Num(3)	3.	Circulatory system disease	1= Occurred in baseline 0= Did not occur	Renamed from Covar5 (CSD)
Num_PVT	Num(3)	3.	Paroxysmal ventricular tachycardia	1= Occurred in baseline 0= Did not occur	Renamed from Covar6 (DEFRIB)
Num_Diabetes	Num(3)	3.	Diabetes mellitus	1= Occurred in baseline 0= Did not occur	Renamed from Covar7 (DM)
Num_Heart_Failure	Num(3)	3.	Heart failure / cardiomyopathy	1= Occurred in baseline 0= Did not occur	Renamed from Covar8 (HF)
Num_Hyperlipidemia	Num(3)	3.	Hyperlipidemia	1= Occurred in baseline 0= Did not occur	Renamed from Covar9 (HPL)
Num_Hypertension	Num(3)	3.	Hypertension	1= Occurred in baseline 0= Did not occur	Renamed from Covar10 (HT)
Num_Kidney	Num(3)	3.	Kidney disease	1= Occurred in baseline 0= Did not occur	Renamed from Covar11 (KID)

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
Num_Malaria	Num(3)	3.	Malaria disease	1= Occurred in baseline 0= Did not occur	Renamed from Covar12 (MALDX)
Num_Malaria_Drug	Num(3)	3.	Dextromethorphan Hbr/quinidine sulfate	1= Occurred in baseline 0= Did not occur	Renamed from Covar13 (MALRX)
Num_Obesity	Num(3)	3.	Obesity	1= Occurred in baseline 0= Did not occur	Renamed from Covar14 (OBES)
Num_Other_Dysrhythmia	Num(3)	3.	Other specified or unspecified cardiac dysrhythmia	1= Occurred in baseline 0= Did not occur	Renamed from Covar15 (OUDYS)
Num_Prem_Beat	Num(3)	3.	Premature beats	1= Occurred in baseline 0= Did not occur	Renamed from Covar16 (PREBEAT)
Num_PST	Num(3)	3.	Paroxysmal supraventricular tachycardia	1= Occurred in baseline 0= Did not occur	Renamed from Covar17 (PSVT)
Num_Seizures	Num(3)	3.	Seizure disorder	1= Occurred in baseline 0= Did not occur	Renamed from Covar18 (SEIZE)
Num_Smoking	Num(3)	3.	Smoking	1= Occurred in baseline 0= Did not occur	Renamed from Covar19 (SMOKE)
Num_Unsp_PT	Num(3)	3.	Unspecified paroxysmal tachycardia	1= Occurred in baseline 0= Did not occur	Renamed from Covar20 (UPAT)
Num_VF_Flutter	Num(3)	3.	Ventricular fibrillation and flutter	1= Occurred in baseline 0= Did not occur	Renamed from Covar21 (VENT)
Torsades_Risk	Char(1)	1.	Risk of Torsades de Pointes (TdP), per CredibleMeds	K=Known P=Possible Risk C=Conditional Risk N=Not classified/unknown B=Baseline cohort default value	Risk value per drug
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	

VariableName	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DisenrollDate	Num(4)	mmddyy10.	Date of last Enr_End after index date through episode end + 365 days	Actual date or .E for no disenrollment	Last date of disenrollment observed from index date through exposure episode end date +365 days. If not disenrolled, set to special missing value .E (enrolled) Allow gap < 45 days between enrollment periods, looking for 365 days at end of episode
DisenrollDays	Num(3)	3.	Days from episode end date to last disenrollment	Negative, 0, positive integers, or .E missing	Allows disenrollment to occur anytime from within episode (negative values and zero) to after episode end (positive values); if not disenrolled, set to special missing value .E
Disenroll_Max_Ind	Char(1)	1.	SDD data end flag	W= Within SDD DP MaxDate P=Past DP MaxDate	Indicator that looking for disenrollment 365 days after episode requires looking past DP MaxDate
FinalEnrollmentStatus	Char(1)	$1.	Final enrollment status	E=Enrolled D=Disenrolled	Status observed at end of episode + 365 days
PostEnroll60Utilization	Num(3)	1.	Med or pharm utilization observed 60 days after date of last disenrollment	1=Utilization 0=No utilization .E=Never disenrolled	Defined as medical or pharmacy utilization from 60 days after disenrollment through end of data availability
PostEnrollUtil_Max_Ind	Char(1)	1.	SDD data end flag	W= Within SDD DP MaxDate P=Past DP MaxDate E=Enrolled/never disenrolled	Indicator that looking for utilization 60 days after disenrollment requires looking past DP MaxDate
AgeMatch	Num(3)	1.	Absolute value of difference of integer age on matched patients	0+	Number of age years difference between matched patients in whole integers; ideal value = 0; set for both Exposed and Baseline cohort rows

II. Aggregate Data Returned to [Lead Analytic Site]From Prior Program

The following is the data dictionary of aggregate data returned to [Lead Analytic Site], produced from Package 1.

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
Level	Char(3)	$3.		000=No stratification 001+=Various levels of combinations of stratification variables	Each level will be for 0, 1, or more stratification variables
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	
DrugName	Char(30)	$30.	Name of index drug	Name of a drug *or* "Multiple"	
DaysSuppliedWeeks	Char(5)	$5.	Days supplied in weeks	0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, 95-99, 100-104, 105+, or "BASE"	TBD: (Drug therapy length)
Year	Num(3)	4.	Year of index	Years within study window: 2000+	Year of index date
Sex	Char(1)	$1.	Sex of patient	M, F	SCDM values of A and U will be combined into O (other)
DisenrollFlag	Char(1)	1.	Disenrollment period flag	E=Disenrolled during episode P=Disenrolled after episode .E=Never disenrolled	Flag variable mapped from DisenrollDays: Negative or zero = E Positive = P .E = E

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DeathSrc	Char(1)	$1.	Source of death date	D=Death table E=Encounter table B=Both tables A=Alive/no death evidence	Requires CIDA modification
Disenroll_Max_Ind	Char(1)	1.	SDD data end flag	W= Within SDD DP MaxDate P=Past DP MaxDate	Indicator that looking for disenrollment 365 days after episode requires looking past DP MaxDate
PostEnroll60Utilization	Num(3)	1.	Med or pharm utilization observed 60 days after date of last disenrollment	1=Utilization 0=No utilization .E=Never disenrolled	Defined as medical or pharmacy utilization from 60 days after disenrollment through end of data availability (i.e., captures changes in utilization patterns)
PostEnrollUtil_Max_Ind	Char(1)	1.	SDD data end flag	W= Within SDD DP MaxDate P=Past DP MaxDate E=Never disenrolled	Indicator that looking for utilization 60 days after disenrollment requires looking past DP MaxDate
Malaria	Num(3)	1.	Malaria	1= Occurred in baseline 0= Did not occur	Renamed from Covar20
Malaria_Rx	Num(3)	1.	Quinidine sulfate	1= Occurred in baseline 0= Did not occur	Renamed from Covar21
Episode_DataEnd	Num(3)	3.	Indicator of scanning DP_MaxDate for episode	0=Did not reach DP_MaxDate for episode end 1=Did reach DP_MaxDate for episode end	Flag to indicate that end of episode overlapped the DP maximum date in the SDD.
Months_To_Next_NDI_Year	Num(3)	2.	# of months	3 ,6 ,9, 12, or .A	# of months to require search of subsequent NDI year, grouped by 3 month intervals
AgeGroup	Char(5)	$5.	Age group	45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-85, 86+	Based on age at index date

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DisenrollDays	Char(7)	$7.	Days from episode end to disenrollment	Groupings, by 30, of negative and positive # of days to disenrollment, or zero. Include a value for day 365 (i.e., 30,60,90,120,150, 180, 210, 240, 270, 300, 330, 360, 361-364, 365, 366-390	Convert patient-level DisenrollDays to grouping by 30 days; e.g., 30, 60, 90, 120,150, 180, 210, 240, 270, 300, 330, 360, 361-364, 365, 366-390, etc.
Reenrollment	Num(3)	1.	Reenrollment observed	1=Re-enrolled ever 0=No reenrollment .E=Never disenrolled	Enr_Start found after disenrollment
FinalEnrollmentStatus	Char(1)	$1.	Final enrollment status	E=Enrolled D=Disenrolled	Status observed at end of episode + 365 days
DeathDays	Char(7)	$7.	Day categories from index date to death	30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360, 361-364, 365, 366-390, >390 or missing value .A	Actual days to be replaced with values up to maximum category No death uses special missing value of A (alive)
DeathYear	Num(3)	4.	Year of death	Years within study window: 2000-2017, or .A (Alive)	Year of death
DeathOccurredFlag	Char(5)	$5.	Time period when death occurred	A=Alive; no death occurred Y_E= Death occurred from IndexDt through EpisodeEndDt Y_365=Death occurred from EpisodeEndDt to EpisodeEndDt + 365 N=Death occurred within some other time interval	Flag to indicate interval of death, if any

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DisenrollYear	Num (3)	4.	Year of disenrollment	Years within study window: 2000-2017 or .N for not disenrolled	Year of disenrollment
Reason_Select	Char(2)	$2.	Reason record selected	DE=Death DI=Disenrolled B=Both NS=Not selected	Reason selected: death, disenrollment, both, or not selected
Count	Num(8)	comma16.	Count within stratification	0+	

III. Aggregate Descriptives Display

Variable	Cohort 1	Cohort 2
Demographics		
Male		
45-54 years	N (% of all males within cohort)	N (% of all males within cohort)
55-64 years	N (% of all males within cohort)	N (% of all males within cohort)
65-74 years	N (% of all males within cohort)	N (% of all males within cohort)
75-85 years	N (% of all males within cohort)	N (% of all males within cohort)
86+ years	N (% of all males within cohort)	N (% of all males within cohort)
Female		
55-64 years	N (% of all females within cohort)	N (% of all females within cohort)
65-74 years	N (% of all females within cohort)	N (% of all females within cohort)
75-85 years	N (% of all females within cohort)	N (% of all females within cohort)
86+ years	N (% of all females within cohort)	N (% of all females within cohort)
Age – Overall		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Age – Male		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Age – Female		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Sex, % female	%	%
Healthcare utilization intensity measures during the baseline period		
Hospitalizations (NumVisits_IP)		
Mean	99.9	99.9

Variable	Cohort 1	Cohort 2
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Emergency department visits (NumVisits_ED)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Ambulatory care visits (NumVisits_AV)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Institutional visits (NumVisits_IS)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Other ambulatory care visits (NumVisits_OA)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Unique medications dispensed (RawDisp)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	999	999
Comorbid conditions, identified during the baseline period		
Arrhythmia / conduction disorder, by type (Num_Arrhythmia_Any)	N (% of cohort)	N (% of cohort)
Atrial fibrillation and flutter (Num_Afib_Flutter)	N (% of cohort)	N (% of cohort)
Paroxysmal ventricular tachycardia (Num_PVT)	N (% of cohort)	N (% of cohort)

Variable	Cohort 1	Cohort 2
Ventricular fibrillation and flutter (Num_VF_Flutter)	N (% of cohort)	N (% of cohort)
Paroxysmal supraventricular tachycardia (Num_PST)	N (% of cohort)	N (% of cohort)
Unspecified paroxysmal tachycardia (Num_Unsp_PT)	N (% of cohort)	N (% of cohort)
Premature beats (Num_Prem_Beat)	N (% of cohort)	N (% of cohort)
Other specified or unspecified cardiac dysrhythmia (Num_Other_Dysrhythmia)	N (% of cohort)	N (% of cohort)
Arrhythmia / conduction disorder, by type (Num_Arrhythmia_Any)	N (% of cohort)	N (% of cohort)
Atrial fibrillation and flutter (Num_Afib_Flutter)	N (% of cohort)	N (% of cohort)
Paroxysmal ventricular tachycardia (Num_PVT)	N (% of cohort)	N (% of cohort)
Ventricular fibrillation and flutter (Num_VF_Flutter)	N (% of cohort)	N (% of cohort)
Cerebrovascular disease (Num_Cerebro)	N (% of cohort)	N (% of cohort)
Coronary heart disease (Num_CHD)	N (% of cohort)	N (% of cohort)
Diabetes mellitus (Num_Diabetes)	N (% of cohort)	N (% of cohort)
Heart failure / cardiomyopathy (Num_Heart_Failure)	N (% of cohort)	N (% of cohort)
Cardioverter-defibrillator/pacemaker (Num_Defrib)	N (% of cohort)	N (% of cohort)
Hyperlipidemia (Num_Hyperlipidemia)	N (% of cohort)	N (% of cohort)
Hypertension (Num_Hypertension	N (% of cohort)	N (% of cohort)
Kidney disease (Num_Kidney)	N (% of cohort)	N (% of cohort)
Circulatory system disease (Num_Circulatory)	N (% of cohort)	N (% of cohort)
Seizure disorder (Num_Seizures)	N (% of cohort)	N (% of cohort)
Smoking (Num_Smoking)	N (% of cohort)	N (% of cohort)
Obesity (Num_Obesity)	N (% of cohort)	N (% of cohort)
Combined Charlson and Elixhauser co-morbidity score (Combined_Score_Num)		
Mean	99.9	99.9
Standard deviation	99.9	99.9
Median	99	99
Min	99	99
Max	99	99
Co-morbidity score (Combined_Score)		
Low-0	N (% of cohort)	N (% of cohort)
1	N (% of cohort)	N (% of cohort)
2-3	N (% of cohort)	N (% of cohort)
4-7	N (% of cohort)	N (% of cohort)
8+	N (% of cohort)	N (% of cohort)
Risk of Torsades de pointes (TdP), per CredibleMeds (Torsades_Risk)		
Known Risk (K)	N (% of cohort)	N/A
Possible risk (P)	N (% of cohort)	N/A
Conditional Risk (C)	N (% of cohort)	N/A
Not classified (N)	N (% of cohort)	N/A

Variable	Cohort 1	Cohort 2
Cohort entry year (Year of Index Date)		
2000	N (% of cohort)	N (% of cohort)
2001	N (% of cohort)	N (% of cohort)
2002	N (% of cohort)	N (% of cohort)
2003	N (% of cohort)	N (% of cohort)
2004	N (% of cohort)	N (% of cohort)
2005	N (% of cohort)	N (% of cohort)
2006	N (% of cohort)	N (% of cohort)
2007	N (% of cohort)	N (% of cohort)
2008	N (% of cohort)	N (% of cohort)
2009	N (% of cohort)	N (% of cohort)
2010	N (% of cohort)	N (% of cohort)
2011	N (% of cohort)	N (% of cohort)
2012	N (% of cohort)	N (% of cohort)
2013	N (% of cohort)	N (% of cohort)
2014	N (% of cohort)	N (% of cohort)
2015	N (% of cohort)	N (% of cohort)
2016	N (% of cohort)	N (% of cohort)
2017	N (% of cohort)	N (% of cohort)

IV. Case Filtering and Preparation for Identifiers

1. [Lead Analytic Site] will determine filtering criteria using the variables from the table above. Each criterion will be "ANDed" with all other criteria when the filtering criteria are applied. A criterion is defined by a variable, an operator and value(s). In some cases a criterion may consist of multiple sub-criteria which are "ORed" together. The following is a description of the filtering criteria elements.

Variable	Char(32)	$32.	Variable	Must be one of the variables from the table in Section I	
Operator	Char(10)	$10.	Operator	Operator Meaning = Equals GT Greater than LT Less than GE Greater than or equal to LE Less than or equal to NE Not equal to IN In (a list of values)	
Values	Char(100)	$100.	Values for filtering	Free text	
Logic	Char(3)	$3.	Logic	OR AND Blank	All criteria rows are assumed "AND" to each other unless noted in combinations

Parentheses will be used to group conditions together as appropriate.

2. An example follows:

Criterion	Variable	Type	Operator	Values	Logic	CritRef	Notes/Comments
1	Year	Num	GE	2010			Use data only where year >= 2010
2	Disenroll_Max_Ind	Char	=	W			Use data only where the disenrollment maximum date occurred with the DP MaxDate
3	Malaria	Num	=	0	AND		Malaria did not occur
4	Malaria_RX	Num	=	0	AND		Use of quinidine sulfate drugs did not occur+
5	Sex	Char	=	M	AND	6	

Criterion	Variable	Type	Operator	Values	Logic	CritRef	Notes/Comments
6	AgeGroup	Char	IN	55-59, 60-64, 65-69, 70-74, 75-79, 80-85, 86+	AND	5	
7	Sex	Char	=	F	AND	8	
8	AgeGroup	Char	IN	60-64, 65-69, 70-74, 75-79, 80-85	AND	7	

(Year GE 2010

AND

Disenroll_Max_Ind = W

AND

Malaria = 0

AND

MalariaRx = 0

AND

((Sex = M AND AgeGroup IN (55-59, 60-64, 65-69, 70-74, 75-79, 80-85, 86+))

OR

(Sex = F AND AgeGroup IN (60-64, 65-69, 70-74, 75-79, 80-85))

OR

(DeathDt between IndexDt and (EpisodeEndDt + 365))

3. The filtering criteria are applied to the patient-level file that had been saved in the DPLocal data library, named **All_Cohorts.** After filtering, save the resulting rows from the patient-level table to a file following the data dictionary below.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
PatID	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.	SOC Program
SSN	Char(9)	$9.	Social Security Number	9 numeric digits, blank	Social Security Number, exactly 9 digits, leading zero filled, to be provided by Data Partner.	DP: Fill with only valid SSNs; digits only
Last_Name	Char(20)	$20.	Last name of patient	Alpha chars, blank, <=1 of ‘ and/or -)	Last Name of Patient to be provided by the Data Partner.	DP: Fill only with valid names, hyphen and/or apostrophe; no extraneous characters
First_Name	Char(15)	$15.	First name of patient	Alpha chars, blank	First Name of Patient to be provided by Data Partner.	DP: Fill only with valid names, hyphen and/or apostrophe; no extraneous characters
Middle_Initial	Char(1)	$1.	Middle initial of patient name	Alpha chars, blank	Patient’s middle initial to be provided by Data Partner.	DP: Fill only with valid letters; no extraneous characters
Month_of_Birth	Num(3)	Z2.	Month of birth	01-12	Patient’s month of birth, with leading zero, automatically filled from SDD.	SOC Program
Day_of_Birth	Num(3)	Z2.	Day of birth	01-31	Patient’s day of birth, with leading zero, automatically filled from SDD.	SOC Program
Year_of_Birth	Num(3)	Z4.	Year of birth	1850-20xx	Patient’s year of birth, automatically filled from SDD.	SOC Program

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
Fathers_Surname	Char(18)	$18.	Fathers surname	Alpha chars, blank, <=1 of ‘ and/or -)	Patient’s father’s surname to be provided by Data Partner.	DP: Fill only with valid names, hyphen and/or apostrophe; no extraneous characters
Age_Unit	Num(3)	1.	Age unit recorded (at death)	0-6, 9, blank	Unit the age is recorded in, unknown and will not be populated (left blank).	SOC Program
Number_Age_Units	Num(3)	Z2.	Number of age units (at death)	00-99, blank	Number of age units at time of death, unknown and will not be populated (left blank).	SOC Program
Sex	Char(1)	$1.	Sex of patient	M, F	Sex of patient, automatically recoded and filled from SDD.	SOC Program
Race	Char(1)	$1.	Race of patient	0-9, blank	Race of patient, automatically recoded and filled from SDD.	SOC Program
Marital_status	Char(1)	$1.	Marital status of patient	1=Never married/single 2=Married 3=Widowed 4=Divorced 5, blank=Unknown	Patient’s marital status to be provided by Data Partner	DP: Fill only with valid values
State_of_Residence	Char(2)	$2.	State of residence	See Appendix A	Patient’s state of residence automatically recoded and filled with last known value in SDD	SOC Program DP: If blank from SOC Program, fill with valid character or numeric values from Appendix A

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
State_of_birth	Char(2)	$2.	State of birth	See Appendix A	State where patient was born, unknown and will not be populated (left blank).	DP: Fill only with valid character or numeric values from Appendix A
Control_Number	Char(10)	$10.	Control number	0000000001+	Sequential number, leading zeros, incremented for each PatID written to this table, inserted automatically	SOC Program
Year_NDI	Num(3)	4.	Year to be searched	2000-2017	NDI year to be searched	SOC Program
Reason_Select	Char(2)	$2.	Reason record selected	DE=Death DI=Disenrolled B=Both	Reason selected: death, disenrollment, or both	SOC Program

4. Copy the following variables:
   4.1. PatID
   4.2. Sex
   4.3. Race
   4.4. State, renamed as State_Of_Residence. Any values not found in Appendix A, shall be set to "99" for unknown.
   4.5. Reason_Select
5. Create the following variables:
   5.1. Month_of_Birth: Month component of Birth_Date.
   5.2. Day_of_Birth: Day component of Birth_Date.
   5.3. Year_of_Birth: Year component of Birth_Date.
   5.4. Year_NDI: Fill as follows:
      5.4.1.If DeathDt is filled, then fill as year portion of DeathDt
      5.4.2.If DeathDt is not filled, then fill with year portion of DisenrollDate.
6. Sort the file by PatID.

7. Set Control_Number as a character variable, per the data dictionary, starting at 1 and incrementing for each unique PatID.
8. Name the file **Patients_For_NDI_Pre_PHI** and save the file to the DPLocal data library.
9. Note that the following variables will be null at this point in the process, for populating by HPs from their source data:
   9.1. SSN
   9.2. Last_Name
   9.3. First_Name
   9.4. Middle_Initial
   9.5. Fathers_Surname
   9.6. Marital_status
   9.7. State_of_birth
   9.8. Age_Unit
   9.9. Number_Age_Units
10. When the file is populated with identifiers, Health Plans will be instructed to save the file to the DPLocal data library and name the file **Patients_For_NDI_PHI_Filled.**
11. In addition to filling the variables named in step 9 above, HPs will be instructed that they can create multiple records, per PatID, when they have more than one name for a patient. When they create multiple records, they will be instructed to keep all variables identical except for the different name information.

V. Pre PHI Aggregate File for [Lead Analytic Site]

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
HP	Char(6)	$6.	Health plan		
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	CohortInd
Reason_Select	Char(2)	$2.	Reason record selected	DE=Death DI=Disenrolled B=Both	Reason selected: death, disenrollment, or both
Year_NDI	Num(3)	4.	Year to be searched	2000-2017	
TotalCount	Num(8)	comma12.	Total count	1+	Total of patients filtered for populating with PHI

Methods to Create the Pre PHI Aggregate file for [Lead Analytic Site]

1. Fill the variable HP with the value in variable DPID_SiteID.
2. Aggregate the **Patients_For_NDI_Pre_PHI**, saved to DPLocal. Aggregate by CohortInd, Reason_Select, and Year_NDI. Save the resulting counts to the TotalCount variable.
3. Save as file **Pre_PHI_Counts** to the MSOC data library.

VI. Aggregate Descriptive Statistics by Cohort - Post Filtering

Using the **Patients_For_NDI_Pre_PHI** file created in NDI Package 1, this aggregate table will be returned to [Lead Analytic Site]. Reference the following from NDI Package 1 specifications:

- Aggregate Descriptive Statistics by Cohort contains the data dictionary and methods for calculations.
- Aggregate Descriptive Display displays how this data may be transposed for display purposes and reference.

Methods to Create the Pre PHI Aggregate file for [Lead Analytic Site]

1. After filtering in Section II, create a temporary table as a subset of the **All_Cohorts** file, saved in the DPLocal data library from package 1. The subset for this table is for only those PatIDs from the **All_Cohorts** file that are found in the **Patients_For_NDI_Pre_PHI** file.
2. Using this temporary file, execute the logic from NDI Package 1, Section IV.
3. Name the resulting file as **NDI_Cohort_Descriptives_Filtered** and save to MSOC data library for this package.

Appendix A – State Postal Codes Accepted by NDI

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Alabama	AL	01	New Jersey	NJ	31
Alaska	AK	02	New Mexico	NM	32
Arizona	AZ	03	New York	NY	33
Arkansas	AR	04	North Carolina	NC	34
California	CA	05	North Dakota	ND	35
Colorado	CO	06	Ohio	OH	36
Connecticut	CT	07	Oklahoma	OK	37
Delaware	DE	08	Oregon	OR	38
District of Columbia	DC	09	Pennsylvania	PA	39
Florida	FL	10	Rhode Island	RI	40
Georgia	GA	11	South Carolina	SC	41
Hawaii	HI	12	South Dakota	SD	42
Idaho	ID	13	Tennessee	TN	43
Illinois	IL	14	Texas	TX	44
Indiana	IN	15	Utah	UT	45
Iowa	IA	16	Vermont	VT	46
Kansas	KS	17	Virginia	VA	47
Kentucky	KY	18	Washington	WA	48
Louisiana	LA	19	West Virginia	WV	49
Maine	ME	20	Wisconsin	WI	50
Maryland	MD	21	Wyoming	WY	51
Massachusetts	MA	22	Puerto Rico	PR	52
Michigan	MI	23	Virgin Islands	VI	53
Minnesota	MN	24	Guam	GU	54
Mississippi	MS	25	Canada	CN	55
Missouri	MO	26	Cuba	CU	56
Montana	MT	27	Mexico	MX	57
Nebraska	NE	28	Remainder of World	RW	59
Nevada	NV	29	Unknown		99
New Hampshire	NH	30			

Program Package 3: Manipulation and Quality Assurance of Records for NDI

I. Populated Records for NDI

This Program Package quality checks files prepared for submission to the NDI. The following is the data dictionary of patient level data to be checked before submission of records to the NDI.

This file will be found in the DPLocal data library for this Program Package 3 at the Health Plan and will be named **Patients_For_NDI_PHI_Filled.**

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Full Sentinel DP identifier	MS Program
PatID	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.	MS Program
SSN	Char(9)	$9.	Social Security Number	9 numeric digits, blank	Social Security Number, exactly 9 digits, leading zero filled, to be provided by Data Partner.	DP
Last_Name	Char(20)	$20.	Last name of patient	Alpha chars, blank, <=1 of ‘ and/or -)	Last Name of Patient to be provided by the Data Partner.	DP
First_Name	Char(15)	$15.	First name of patient	Alpha chars, blank	First Name of Patient to be provided by Data Partner.	DP
Middle_Initial	Char(1)	$1.	Middle initial of patient name	Alpha chars, blank	Patient’s middle initial to be provided by Data Partner.	DP
Month_of_Birth	Num(3)	Z2.	Month of birth	01-12, blank	Patient’s month of birth, with leading zero, automatically filled from SDD.	MS Program
Day_of_Birth	Num(3)	Z2.	Day of birth	01-31, blank	Patient’s day of birth, with leading zero, automatically filled from SDD.	MS Program

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
Year_of_Birth	Num(3)	Z4.	Year of birth	1850-20xx, blank	Patient's year of birth, automatically filled from SDD.	MS Program
Fathers_Surname	Char(18)	$18.	Fathers surname	Alpha chars, blank, <=1 of ' and/or -)	Patient's father's surname to be provided by Data Partner.	DP
Age_Unit	Num(3)	1.	Age unit recorded (at death)	0-6, 9, blank	Unit the age is recorded in, unknown and will not be populated (left blank).	MS Program
Number_Age_Units	Num(3)	Z2.	Number of age units (at death)	00-99, blank	Number of age units at time of death, unknown and will not be populated (left blank).	MS Program
Sex	Char(1)	$1.	Sex of patient	M, F, blank	Sex of patient, automatically recoded and filled from SDD.	MS Program
Race	Char(1)	$1.	Race of patient	0-8, blank	Race of patient, automatically recoded and filled from SDD.	MS Program
Marital_status	Char(1)	$1.	Marital status of patient	1-4, blank	Patient's marital status to be provided by Data Partner	DP
States_of_Residence	Char(2)	$2.	State of residence	see Appendix B	Patient's state of residence automatically recoded and filled with last known value in SDD[1].	MS Program
State_of_birth	Char(2)	$2.	State of birth	See Appendix B	State where patient was born, unknown and will not be populated (left blank).	DP
Control_Number	Char(10)	$10.	Control number	0000000001+	Sequential number, leading zeros, incremented for each PatID written to this table, inserted automatically	MS Program

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments	Entity Source
Sequence_Number	Char(2)	$2.	Sequence number	01+	Sequential number, leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"	MS Program
NDI_Eligible	Num(3)	1.	Eligible for NDI matching	1=Eligible 0=Eligible	Based on error FlagID Eligible_1 from Appendix A	MS Program

Methods to create the patient-level file for Health Plans

1. Check that the file named **Patients_For_NDI_PHI_Filled** is found in the DPLocal data library for this programming Package 3.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library.
    b. Otherwise if found, proceed with all other processes.
2. Check that the file named **R01_Signature** is found in the MSOC data library from Package 1.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library.
    b. Otherwise if found, proceed as follows:
        i. Scan for the row where Var = "DPMaxDate".
        ii. Create a temporary variable (e.g., macro variable), DPMaxDate, and set it to the date value of the character value found in the Value variable. This will be used in Section IV below
        iii. Continue with step 6 in this section.
3. Determine if the health plan identifies a file that lists PatIDs not eligible for research.
    a. If so, then on PatID, merge **Patients_For_NDI_PHI_Filled** with such file, keeping all rows from **Patients_For_NDI_PHI_Filled** with PatIDs not found in the health plan file that lists PatIDs not eligible for research. Save this file temporarily as **Patients_For_NDI_PHI_Filled_Elig**; continue with the next step.
    b. Otherwise, copy **Patients_For_NDI_PHI_Filled** to temporary file **Patients_For_NDI_PHI_Filled_Elig** and continue with step 7.
4. Check that the file named **Patients_For_NDI_Pre_PHI** is found in the DPLocal data library from Package 2.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 2.

b. Otherwise if found, proceed with next steps.

5. Link temporary file with saved Demographics file: Link **Patients_For_NDI_PHI_Filled_Elig** with **Patients_For_NDI_Pre_PHI** found in the DPLocal data library from Package 2 on PatID, in order to obtain Birth_Date, Sex, Race, and State_of_Residence variables.
    a. If the link fails, then produce the following outcomes:
        i. Write a dataset to the DPLocal data library, named **Patients_NDI_Link_Failure_Dem** which contain one variable, PatID. Insert into this file those PatIDs found in **Patients_For_NDI_PHI_Filled_Elig** but not in **Patients_For_NDI_Pre_PHI**.
        ii. Save the number of records in **Patients_NDI_Link_Failure_Dem** to a temporary variable.
        iii. Write an error message to the log stating:
            1. # PatIDs in **Patients_For_NDI_PHI_Filled** were not found in the **Demographic_NDI** file. ("#" shall be replaced with the value from step 8.a.ii.)
            2. Identity of the failed PatIDs can be found in the DPLocal file **Patients_NDI_Link_Failure_Dem**.
        iv. Abort the process.
        v. NOTE: The message in step 8.a.iii.1 references the **Patients_For_NDI_PHI_Filled** file, to which health plans have access instead of the temporary file **Patients_For_NDI_PHI_Filled_Elig,** which will not be available to health plans if this program is aborted.
    b. If the link succeeds, then perform the next steps.
6. Check consistency of variables that come from the SCDM Demographics table:
    a. Combine Month_of_Birth, Day_of_Birth, and Year_of_Birth from **Patients_For_NDI_PHI_Filled_Elig** into a Birth_Date variable.
    b. Check that Birth_Date, Sex, Race, and State_Of_Residence from **Patients_For_NDI_PHI_Filled_Elig** match the values in the **Demographic_NDI** file.
    c. If any of the 4 variables do not match, then perform the following:
        i. Write a file to the DPLocal data library for this Package 3, called **Pats_Changed_Dem**, that contains rows for only those PatIDs with any changed values. The variables in this file are: PatID, Birth_Date, Sex, Race, and State_Of_Residence. Fill these variables with the values from the **Patients_For_NDI_PHI_Filled_Elig** file.
        ii. Save the number of records in **Pats_Changed_Dem** to a temporary variable.
        iii. Write a warning message to the log stating:
            1. # PatIDs in **Patients_For_NDI_PHI_Filled** contain differences in values from those in the original ETL used for Package 1. ("#" shall be replaced with the value from step 9.c.ii.)
            2. Identity of the PatIDs with different values can be found in the DPLocal file **Pats_Changed_Dem**.
        iv. Proceed with next steps.
7. For each PatID, process the Last_Name variable.

a. Drop name generational suffixes: These are any of the following case insignificant strings: Jr, Sr, I, II, III, IV, V, MD, PhD, with or without periods. Drop the periods if they are part of the generational suffix.
b. Combine name components by removing delimiters: Remove single quote and single space delimiters only for the following:
    i. Single letters of D or O, as in O'Leary or D'Santis; combine as DSantis or OLeary
    ii. The following words: Ben, Da, De, Del, Den, Der, La, Le, Mac, Mc, San, St, Ste, Van. Example: Van Heusen becomes VanHeusen; Le Duan becomes LeDuan; Van De Lay becomes VanDeLay.

8. Create additional rows: If there are two or more name strings within this Last_Name variable, defined as name strings separated by a space or a hyphen, for each name string:
    a. Create a new row.
    b. Fill the row with the same variable values from the examined row, with the exception of Last_Name.
    c. Fill Last_Name with the specific name string.
    d. Remove spaces and hyphens on the added rows.

    Examples:

    Sylvia Mathews Burwell becomes a total of three rows:

    - Row #1 of 3: First_Name=Sylvia, Last_Name= Mathews Burwell, as the original row
    - Row #2 of 3: First_Name=Sylvia, Last_Name=Mathews, as an added row
    - Row #3 of 3: First_Name=Sylvia, Last_Name=Burwell, as an added row

    Maria Contreras-Sweet becomes two rows:

    - Row #1 of 3: First_Name=Maria, Last_Name= Contrera-Sweet, as the original row
    - Row #2 of 3: First_Name=Maria, Last_Name=Contreras
    - Row #3 of 3: First_Name=Maria, Last_Name=Sweet
9. Create the Sequence_Number variable. This indicates how many name records per PatID. For each PatID, start at 1 and increment by 1 for each subsequent row with the same PatID. Set as a character variable with a leading zero, if needed. Each PatID will have at least one row with a value of "01".
10. Create a placeholder variable NDI_Eligible and fill all values with missing. This variable will be filled in the next section.
11. Save this file as **Patients_For_NDI_Filled_PHI2** in the DPLocal data library.

II. Health Plan Patient-Level Report

The following is a patient-level report dataset, to be provided to the Health Plan ONLY, with indicators of quality control issues.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Patient identifier
FlagID	Char(21)	$21.	Flag identifier	Free text	Code for error identified
Flag_Descr	Char(255)	$255.	Flag description	Free text	Description of error identified
Message	Char(255)	$255.	Error message	Free text	Details for error

Methods to create the patient-level file for Health Plans

1. Review each row in the DPLocal file **Patients_For_NDI_Filled_PHI2**. For each issue identified, using Appendix A, create a row in the table described above. NOTE: This checking should occur only for those variables that are filled. If a variable is blank, as noted in the data dictionary in Section I, then no checking is needed, as these blanks are permitted in records sent to NDI.
    a. Fill the PatID with the PatID from the row being examined.
    b. Fill the FlagID with the value from the Appendix A table.
    c. Fill the Flag_Descr with the value from the Appendix A table.
    d. Fill the Message with the value from the row being examined, for the variable(s) and value(s) examined.
2. Name the file **Patients_For_NDI_Errors**, sort by PatID, and save to the DPLocal data library.
3. In the **Patients_For_NDI_Filled_PHI2** file created in the prior section and saved to the DPLocal data library, fill variable NDI_Eligible as follows:
    a. If PatID is found in **Patients_For_NDI_Errors** and has a FlagID="Eligible_1", then set to 0.
    b. If PatID is not found in **Patients_For_NDI_Errors** with a FlagID="Eligible_1", then set to 1.

III. Lead Analytic Site Aggregate Level Error Report

The following is an aggregate dataset, to be provided to [Lead Analytic Site], with counts of quality control issues.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
FlagID	Char(21)	$21.	Flag identifier	Free text	Code for error identified
Flag_Descr	Char(255)	$255.	Flag description	Free text	Description of error identified
Count	Num(8)	comma12.	Count	1+	Count for error identified

Methods to create this aggregate file for Health Plans

1. Using the file **Patients_For_NDI_Errors**, aggregate the file by FlagID.
    a. Fill DPID_SiteID.
    b. Fill FlagID with the FlagID in the aggregate file.
    c. Fill Flag_Descr with the corresponding Flag_Descr variable.
    d. Fill Count with the counts calculated.
2. Sort the aggregate file by FlagID, name the file **DPID_SiteID_NDI_Errors**, and save to the MSOC data library.

IV. Lead Analytic Site Aggregate Level Counts Report

The following is an aggregate dataset, to be provided to [Lead Analytic Site], with counts of number of records by NDI year to be searched.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
FileInd	Char(5)	$5.	File indicator	Pre = Patients_For_NDI_Pre_PHI Post = Patients_For_NDI_Filled_PHI2	Indicator of file used for counts; "post" may have had filtering for patients no longer eligible for research
Year_NDI	Num(3)	4.	Year	2000+	NDI year to be searched
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Count	Num(8)	comma12.	Count	1+	Count for year

Methods to create this aggregate file for [list Lead analytic site]

1. Check that the file named **All_Cohorts** is found in the DPLocal data library for Package 1.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 1.
    b. Otherwise if found, proceed with subsequent steps.
2. Using the file **Patients_For_NDI_Filled_PHI2**, link with file **All_Cohorts** (found in the Package 1 DPLocal data library), on PatID, in order to obtain variables CohortInd, DeathDt DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind. Keep all rows from the file **Patients_For_NDI_Filled_PHI2**. If the link fails, then produce the following outcomes:
    a. Write a dataset to the DPLocal data library, named **Patients_For_NDI_Link_Failure** which shall contain only one variable, PatID. Insert into this file those PatIDs found in **Patients_For_NDI_Filled_PHI2** but not in **All_Cohorts**.
    b. Write an error message to the log stating:
        i. PatIDs in **Patients_For_NDI_Filled_PHI2** were not found in the **All_Cohorts** file.

ii. Identity of the failed PatIDs can be found in the DPLocal file **Patients_For_NDI_Link_Failure**.

c. Abort the process.

3. If the link succeeds, then perform the next steps.
4. Check that the file named **Patients_For_NDI_Pre_PHI** is found in the DPLocal data library for Package 2.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 2.
    b. Otherwise if found, proceed with subsequent steps.
5. Link with file **All_Cohorts** (found in the Package 1 DPLocal data library), on PatID, in order to obtain variables CohortInd, DeathDt DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind. Keep all rows from the file **Patients_For_NDI_Pre_PHI**. NOTE: There is no need in this package to check that this link succeeds, as in steps 2.a-2.c above, as the **Patients_For_NDI_Pre_PHI** file was written under package 2.
6. Fill DPID_SiteID.
    a. Concatenate the file that results from step 3 above (i.e., SQL union) with the **Patients_For_NDI_Pre_PHI** file, keeping variables PatID, CohortInd, DeathDt DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind.
    b. Create variable FileInd as follows:
        i. If the row is in the temporary file at this stage (i.e., **Patients_For_NDI_Filled_PHI2** (with variables CohortInd, DeathDt, DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind), then set as "Post".
        ii. If the row comes from the **Patients_For_NDI_Pre_PHI** file (with variables CohortInd, DeathDt, DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind), set as "Pre".
    c. Year_NDI: Fill as follows:
        i. If DeathDt is filled, then fill as year portion of DeathDt
        ii. If DeathDt is not filled, then fill with year portion of DisenrollDate.
        iii. If both DeathDt and DisenrollDate are not filled, then check the following: If DeathSrc is not one of: D, E, nor B (i.e., no death) *and* FinalEnrollmentStatus is not D (i.e., still enrolled) *and* Disenroll_Max_Ind equals P (i.e., need to look past DP maximum date to determine disenrollment), then use the year portion of the temporary variable DPMaxDate. NOTE: This will have been obtained in Section I, step 2. Otherwise, do not fill the row with Year_NDI.
    d. Aggregate the resulting file on FileInd, CohortInd, and Year_NDI.
    e. Fill Count with the counts calculated.
7. Sort the aggregate file by FileInd, CohortInd and Year_NDI, name the file **DPID_SiteID_NDI_SubmissionCounts**, and save to the MSOC data library.

Appendix A – Error Flags

FlagID	Flag_Descr	Message	Rules and Examples
PatID_1	PatID not found in patient level file	Fill with PatID	PatID must be found in the patient-level table written to DPLocal by Package 2, **Patients_For_NDI_Pre_PHI**
SSN_1	SSN contains values other than digits	Fill with SSN	09852445A
SSN_2	SSN contains fewer than 9 digits	Fill with SSN	09852
SSN_3	Invalid SSN (see Appendix C)	Fill with SSN	123456789
Last_Name_1	Last_Name contains values other than letters, " " (space), "-" (hyphen) and "'" (apostrophe)	Fill with Last_Name	Smith8
First_Name_1	First_Name contains values other than letters, " " (space), "-" (hyphen) and "'" (apostrophe)	Fill with First_Name	Ann^
Middle_Initial_1	Middle_Initial contains values other than letters	Fill with Middle_Initial	$
Month_of_Birth_1	Month_of_Birth contains values other than 1-12, with leading zero	Fill with Month_of_Birth	13
Day_of_Birth_1	Day_of_Birth contains values other than 1-31, with leading zero	Fill with Day_of_Birth	32
Day_of_Birth_2	Day_of_Birth contains an invalid value for Month_of_Birth; i.e., 29, 30, or 31	Fill with Day_of_Birth	29 when Month_of_Birth=02 and not a leap year 30 or 31 when Month_of_Birth=02 for any year 31 when Month_of_Birth=04, 06, 09, or 11

FlagID	Flag_Descr	Message	Rules and Examples
Year_of_Birth_1	Year_of_Birth contains values other than digits	Fill with Year_of_Birth	19RR
Year_of_Birth_2	Year_of_Birth contains invalid year values	Fill with Year_of_Birth	Must be >=1910
Fathers_Surname_1	Fathers_Surname contains values other than letters, " " (space), "-" (hyphen) and "'" (apostrophe)	Fill with Fathers_Surname	Youman##
Age_Unit_1	Value other than 0-6, 9, or blank	Fill with Age_Unit	7
Number_Age_Units_1	Unlikely value when considered with Age_Unit	Fill with Age_Unit concatenated with Number_Age_Units	188
Sex_1	Value not M, F, or blank	Fill with Sex	U
Race_1	Value not 0-9 or blank	Fill with Race	A
Marital_Status_1	Value not 1-4 or blank	Fill with Marital_Status	8
State_of_Residence_1	Invalid code	Fill with State_of_Residence	MU
State_of_Birth_1	Invalid code	Fill with State_of_Birth	ZZ
Eligible_1	Missing variable combinations to be eligible for matching by NDI	Fill with: First_Name=, Last_Name=, SSN=, Month_of_Birth=, Year_of_Birth=, calculated date of birth=, and Sex=	Each record must meet at least one of the following criteria; it is an error if *none of the* 3 criteria are not met: 1. First and Last name and Social Security Number 2. First and Last name and Month and Year of birth 3. Social Security Number and Full date of birth and Sex

Appendix B – State Postal Codes Accepted by NDI

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Alabama	AL	01	New Jersey	cc	31
Alaska	AK	02	New Mexico	NM	32
Arizona	AZ	03	New York	NY	33
Arkansas	AR	04	North Carolina	NC	34
California	CA	05	North Dakota	ND	35
Colorado	CO	06	Ohio	OH	36
Connecticut	CT	07	Oklahoma	OK	37
Delaware	DE	08	Oregon	OR	38
District of Columbia	DC	09	Pennsylvania	PA	39
Florida	FL	10	Rhode Island	RI	40
Georgia	GA	11	South Carolina	SC	41
Hawaii	HI	12	South Dakota	SD	42
Idaho	ID	13	Tennessee	TN	43
Illinois	IL	14	Texas	TX	44
Indiana	IN	15	Utah	UT	45
Iowa	IA	16	Vermont	VT	46
Kansas	KS	17	Virginia	VA	47
Kentucky	KY	18	Washington	WA	48

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Louisiana	LA	19	West Virginia	WV	49
Maine	ME	20	Wisconsin	WI	50
Maryland	MD	21	Wyoming	WY	51
Massachusetts	MA	22	Puerto Rico	PR	52
Michigan	MI	23	Virgin Islands	VI	53
Minnesota	MN	24	Guam	GU	54
Mississippi	MS	25	Canada	CN	55
Missouri	MO	26	Cuba	CU	56
Montana	MT	27	Mexico	MX	57
Nebraska	NE	28	Remainder of World	RW	59
Nevada	NV	29	Unknown		99
New Hampshire	NH	30			

Appendix C – Invalid SSNs

The following are indications of an invalid SSN:[3]

1. Invalid whole number: Some specific 9-digit SSNs are considered invalid. These are as follows:
    a. All 9 positions with the same digit; e.g., 999999999
    b. Specific number: 078051120[4]
    c. Specific range: Any in the range of 987-65-4320 through 987-65-4329
2. Invalid values in positions 1-3 (area numbers): 000, 666 and any of 900-999.
3. Invalid values in positions 4-5 (group numbers): 00.
4. Invalid values in positions 6-9 (serial numbers): 0000.

[3] https://secure.ssa.gov/poms.nsf/lnx/0110201035

[4] https://www.ssa.gov/history/ssn/misused.html

Program Package 4: Write Outbound Files for NDI

I. Outbound File for NDI

This Program Package creates outbound files in the NDI required format, for submission to the NDI.

The following is the data dictionary of patient level data to be exported to flat text files for transmission to NDI. Variables with indications in the "Column Positions" column are to be exported as character data only. All other variables are not to be exported. This section includes two successive filtering steps of patient-level data, based on the **Patients_For_NDI_Filled_PHI2** file from package 3:

1. Keeping patients identified by the health plan as still eligible for research
2. Keeping patients that have a combination of PHI variables eligible for matching by the NDI

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Not exported	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
DeathDt	Not exported	Num(4)	mmddyy10.	Death date	Valid SAS dates, or special missing .A, if still alive	From Package 1, All_Cohorts file
DisenrollDate	Not exported	Num(4)	mmddyy10.	Date of last Enr_End after index date through episode end + 365 days	Actual date or .E for no disenrollment	From Package 1, All_Cohorts file
FinalEnrollmentStatus	Not exported	Char(1)	$1.	Final enrollment status	E=Enrolled D=Disenrolled	Status observed at end of episode + 365 days

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
DeathSrc	Not exported	Char(1)	$1.	Source of death date	D=Death table E=Encounter table B=Both tables A=Alive/no death evidence	Requires CIDA modification or ad hoc code
Disenroll_Max_Ind	Not exported	Char(1)	1.	SDD data end flag	W= Within SDD DP MaxDate P=Past DP MaxDate	Indicator that looking for disenrollment 365 days after episode requires looking past DP MaxDate
Last_Name	1-20	Char(20)	$20.	Last name of patient	Alpha chars, blank, <=1 of ‘ and/or -)	Left justified: Last Name of Patient to be provided by the Data Partner.
First_Name	21-35	Char(15)	$15.	First name of patient	Alpha chars, blank	Left justified: First Name of Patient to be provided by Data Partner.
Middle_Initial	36	Char(1)	$1.	Middle initial of patient name	Alpha chars, blank	Patient’s middle initial to be provided by Data Partner.
SSN	37-45	Char(9)	$9.	Social Security Number	9 numeric digits, blank	Social Security Number, exactly 9 digits, maintaining leading zeros, to be provided by Data Partner.
Month_of_Birth	46-47	Num(3)	Z2.	Month of birth	01-12, blank	Patient’s month of birth, with leading zero, automatically filled from SDD.
Day_of_Birth	48-49	Num(3)	Z2.	Day of birth	01-31, blank	Patient’s day of birth, with leading zero, automatically filled from SDD.
Year_of_Birth	50-53	Num(3)	4.	Year of birth	1850-20##, blank	Patient’s year of birth, automatically filled from SDD.

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Fathers_Surname	54-71	Char(18)	$18.	Fathers surname	Alpha chars, blank, <=1 of ‘ and/or -)	Left justified: Patient’s father’s surname to be provided by Data Partner.
Age_Unit	72	Num(3)	1.	Age unit recorded (at death)	0= Died < 100 years 1=Died>= 100 years or Blank	Unit the age is recorded in, unknown and will not be populated (left blank).
Number_Age_Units	73-74	Num(3)	Z2.	Number of age units (at death)	00-99, blank	Number of age units at time of death; unknown will not be populated (left blank).
Sex	75	Char(1)	$1.	Sex of patient	M, F	Sex of patient, automatically recoded and filled from SDD.
Race	76	Char(1)	$1.	Race of patient	0-8, blank	Race of patient, automatically recoded and filled from SDD.
Marital_Status	77	Char(1)	$1.	Marital status of patient	1-4, blank	Patient’s marital status to be provided by Data Partner
State_of_Residence	78-79	Char(2)	$2.	State of residence	see Appendix A	Patient’s state of residence automatically recoded and filled with last known value in SDD[1].
State_of_birth	80-81	Char(2)	$2.	State of birth	See Appendix A	State where patient was born, unknown and will not be populated (left blank).
Control_Number	82-91	Char(10)	$10.	Control number	0000000001+	Sequential number, with leading zeros, incremented for each PatID written to this table, inserted automatically

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Sequence_Number	92-93	Char(2)	$2.	Sequence number	01+	Sequential number, with leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"
DPID	94-99	Char(6)	$6.	Health plan identifier	AEOS, HMHPRF, KPCO, KPNW, KPWA, VBOS	Left justified health plan identifier; will be used to confirm that data files are returned to correct health plan

Methods to create the patient-level text file for transmission to the NDI

1. Check that the file named **R01_Signature** is found in the MSOC data library from Package 1.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the MSOC data library.
    b. Otherwise if found, proceed as follows:
        i. Scan for the row where Var = "DPMaxDate".
        ii. Create a temporary variable (e.g., macro variable), DPMaxDate, and set it to the date value of the character value found in the Value variable. This will be used in Sections II and III below.
        iii. Continue with next step in this section.
2. Check that the file named **Patients_For_NDI_Filled_PHI2** is found in the DPLocal data library from Package 3.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from package 3.
    b. Otherwise if found, continue with next step in this section.
3. Use the file **Patients_For_NDI_Filled_PHI2** that had been saved, by Data Partners, in the DPLocal data library from Package 3. Note that this file is partially filled by Data Partners with their source data and is finalized by a SOC program.
4. Create the Sequence_Number variable. This indicates how many records per PatID. For each PatID, start at 1 and increment by 1 for each subsequent row with the same PatID. Set as a character variable with a leading zero, if needed. Each PatID will have at least one row with a value of "01".
5. Determine if the health plan identifies a file that lists PatIDs not eligible for research.

   a. If so, then on PatID, merge **Patients_For_NDI_Filled_PHI2** with such file, keeping all rows from **Patients_For_NDI_Filled_PHI2** containing PatIDs not found in the health plan file that lists PatIDs, not eligible for research. Save this file temporarily as **Patients_For_NDI_Research**; continue with step 6.
   b. Otherwise, copy **Patients_For_NDI_Filled_PHI2** to temporary file **Patients_For_NDI_Research** and continue with step 6.
6. Filter **Patients_For_NDI_Research** for those rows where variable NDI_Eligible=1, to ensure that only patients eligible for matching with the NDI are maintained for the next steps in this section.
7. Link this file, by PatID, to the **All_Cohorts** that had also been saved in the DPLocal data library from Package 1, keep only those rows from the **Patients_For_NDI_Research** file. Add variables DeathDt, DisenrollDate, FinalEnrollmentStatus, DeathSrc, and Disenroll_Max_Ind. These five variables will assist in populating the aggregate file discussed in Sections 0 below.
8. Then save to the DPLocal data library as file named **NDI_Outbound_Records**.
9. Using the temporary file created in step 8, write out a flat text file for transmission to the NDI, with these features:
   a. Fixed column position(s) for data.
   b. No delimiters (i.e., no tabs, no commas).
   c. No header row containing variable names is to be created. All rows should contain data only.
   d. Any SAS missing value (numeric or character) should be written out as a blank
10. Variables are to be filled as noted by the "Column Positions" column above.
11. Name the file as indicated by this table.

Health Plan (Full DP identifier)	**File Name**
Health Plan 1	Health Plan1_Outbound_To_NDI_*yyyymmdd*
Health Plan 2	Health Plan2_Outbound_To_NDI_*yyyymmdd*
Health Plan 3	Health Plan3_Outbound_To_NDI_*yyyymmdd*
Health Plan 4	Health Plan4_Outbound_To_NDI_*yyyymmdd*

   a. The "yyyymmdd" is to be replaced by the value representing the date the program is executed.
   b. The file name is to be followed by a ".TXT" extension.
12. Save the file to the DPLocal data library established for this program package.

II. Aggregate Totals by Year and Health Plan for Lead Analytic Site

An aggregate file is created, for use by health plans and returned to [Lead Analytic Site], to assist the Institute with completing the NDI Transmittal Forms. These transmittal forms indicate totals of records, number of alias records, and range of years to be searched. This is a sample worksheet for the transmittal form:

Years Searched	Number of Years		Number of Subjects		NDI Fee (see above)		NDI Charges (for each subgroup)
1992-2002	11	x	100	x	.21	=	231.00
1993-2002	10	x	57	x	.21	=	119.00
1994-2002	9	x	80	x	.21	=	151.20
1995-2002	8	x	110	x	.21	=	184.80
1996-2002	7	x	65	x	.21	=	95.55

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Full Sentinel DP identifier
Year_NDI_Start	Num(3)	4.	Starting year of search	2000-2017	Starting NDI death year to be searched
Year_NDI_End	Num(3)	4.	Ending year of search	2000-2017	Ending NDI death year to be searched
Number_Years	Num(3)	2.	Number of years	1+	Number of years defined by range of Year_NDI_Start and Year_NDI_End
Study_Subjects	Num(4)	comma12.	Number of patients	0+	Number of distinct patients
NDI_Charges	Num(5)	Comma12.2	Dollar charges	0+	Calculated charges

Methods to create the aggregate file by year for return to [Lead Analytic Site]

1. Establish Parameters: Create parameters as follows that will be set prior to finalizing the program from this specification:
    a. Years Per Record: One or more years may be requested to search per PatID submitted. NOTE: The NDI charges per PatID/patient and number of years to be searched. There is no additional charge for "alias" records; i.e., those with Sequence_Number >= "02". Create three parameters indicating the number of years to search. Each of these will be set to the value of 1 (no duplicate rows), 2 (1 duplicate row for searching one successive year), or 3 (2 duplicate rows for searching one successive year and one prior year). For values 2 and 3, this will result in duplicate rows per PatID; note steps 3 and 4 below.
        i. DeathSelectedYears: This is the number of years to be searched, based on health plan indication that the patient has died, using the patient-level variable DeathDt.
        ii. DisenrollmentSelectedYears: This is the number of years to be searched, based on health plan indication that the patient has disenrolled, using the patient-level variable DisenrollDate.
        iii. LTFSelectedYears: This is the number of years to be searched, based on health plan indication that the patient is lost to followup, as they are still enrolled, but scanning for disenrollment requires looking past DP_MaxDate.
    b. NDI Search Fee: Create a parameter, NDI_Search_Fee, and set to 0.21; i.e., 21 cents.
    c. Latest Year to Search: Create a parameter, LatestSearchYear, and set based on the year portion of temporary variable DPMaxDate. NOTE: This will have been obtained in Section I, step 1
2. Using the file **NDI_Outbound_Records** that was saved in the DPLocal data library in Section I. above , keep variables PatID, Sequence_Number, DeathDt and DisenrollDate. Create the Year_NDI variable as follows:
    a. If DeathDt is filled, then fill with year portion of DeathDt
    b. If DeathDt is not filled, then fill with year portion of DisenrollDate.
    c. If both DeathDt and DisenrollDate are not filled, then check the following: If DeathSrc is not one of: D, E, nor B (i.e., no death) *and* FinalEnrollmentStatus is not D (i.e., still enrolled) *and* Disenroll_Max_Ind equals P (i.e., need to look past DP maximum date to determine disenrollment), then use the year portion of the temporary variable DPMaxDate. NOTE: This will have been obtained in Section I, step 1. Otherwise, do not fill the row with Year_NDI.
    d. QC step: All rows should have a filled Year_NDI value as a result of steps 2.a through 2.c.
3. Read the file resulting from step 2 and create duplicate rows as follows:
    a. **DeathSource:** For records where DeathDt is filled with an actual date, create a number of additional rows defined as DeathSelectedYears minus 1 (e.g., If DeathSelectedYears =2, create 1 additional row per record with DeathDt filled; in this example there will be 2 records per PatID).

b. **DisenrollSource:** For records where DisenrollDate is filled with an actual date, create a number of additional rows defined as DisenrollmentSelectedYears minus 1 (e.g., If DisenrollmentSelectedYears = 3, create 2 additional rows per record with DisenrollDt filled; in this example there will be 3 records per PatID combination).
c. **LostToFollowupSource:** When both DeathDt and DisenrollDate are not filled (i.e., patients lost to follow-up – see step 2.c above), create a number of additional rows defined as LTFSelectedYears minus 1 (e.g., If LTFSelectedYears = 1, create 0 additional rows per record with both DeathDt and DisenrollDt missing; in this example there will be only 1 record per PatID combination).

4. Set Year_NDI for duplicate rows: Follow these rules for setting the value of Year_NDI for duplicate rows, for both DeathSelectedYears and DisenrollmentSelectedYears:
   a. When 2, increment Year_NDI by 1 for the duplicate row; i.e., one later year to be searched. Example: Initial Year_NDI=2014. Duplicate row will have Year_NDI set to 2015.
   b. When 3, create 1 row with Year_NDI incremented by 1 and 1 row with Year_NDI decremented by 1. Example: Initial Year_NDI=2014. Duplicate rows will have one with Year_NDI set to 2015, and one with Year_NDI set to 2013.
   c. DO NOT create any duplicate rows where a resulting Year_NDI will be greater than the value in parameter LatestSearchYear.

   Note that duplicate rows may be created for all source rows per *PatID*, regardless of the value of Sequence_Number. For example, if in the **NDI_Outbound_Records** file, we have a name record as "Maria Contreras-Sweet", where First_Name=Maria and Last_Name=Contreras-Sweet. Package 3 will have created two additional rows in the resulting **NDI_Outbound_Records** file, with Sequence_Number set as follows:

   - Row #1 of 3: First_Name=Maria, Last_Name= Contrera-Sweet, is the original row; Sequence_Number=01.
   - Row #2 of 3: First_Name=Maria, Last_Name=Contreras; Sequence_Number=02.
   - Row #3 of 3: First_Name=Maria, Last_Name=Sweet; Sequence_Number=03.

   Then in this section of package 4, if this record had a DeathDt=5/24/2011 and the DeathSelectedYears parameter=2, then we need create only 1 additional record *per PatID*. The first record for this patient would have Year_NDI=2011 and the second record for this patient would have Year_NDI=2012, based on the rules in step 4.a.

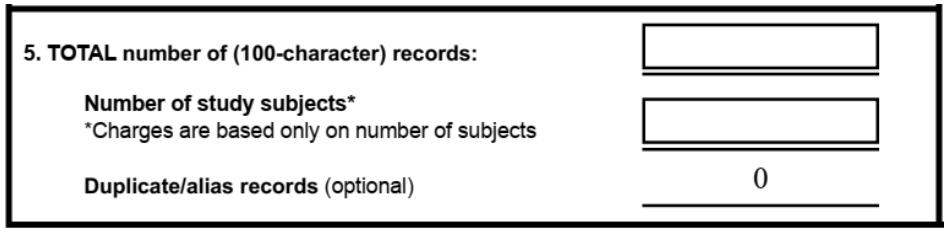
5. Per distinct PatID, fill as follows:
   a. Year_NDI_Start: Earliest Year_NDI
   b. Year_NDI_End: Latest Year_NDI
6. Aggregate the file resulting from the prior step on Year_NDI_Start and Year_NDI_End to get a total count.
   a. Fill Study_Subjects with this count.
   b. For each aggregated row, fill Number_years with Year_NDI_End minus Year_NDI_Start plus 1.

c. Multiply Study_Subjects by Number_Years and by the parameter NDI_Search_Fee, round to 2 decimal places and set NDI_Charges to this value.

7. Fill the data dictionary above, sort by Year_NDI_Start and Year_NDI_End. Save as file ***DPID_SiteID*_Aggregates_Year** to the MSOC data library. DPID_SiteID takes on the value filled in this variable.

III. Aggregate Totals by Health Plan for Lead Analytic Site

An aggregate file is created, for use by health plans and returned to [Lead Analytic Site], to assist the Institute with completing the NDI Transmittal Form and track counts of patients retained during filtering steps. The transmittal form indicates totals of records, number of alias records, and range of years to be searched. This is an excerpt from the transmittal form.



Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Full Sentinel DP identifier
FileInd	Char(5)	$5.	File indicator	PRE1 = Patients_For_NDI_Filled_PHI2 PRE2 = Patients_For_NDI_Research POST = NDI_Outbound_Records	Indicator of file used for counts; PRE2 and POST may have had filtering for patients no longer eligible for research
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Total_Records	Num(4)	comma12.	Number of records	1+	Number of records submitted

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Study_Subjects	Num(4)	comma12.	Number of patients	1+	Number of distinct patients
Alias_Records	Num(4)	comma12.	Number of alias records	0+	Number of alias records; these are the number of *additional* records per patient

Methods to create the aggregate file by year for return to [Lead Analytic Site]

1. Use and concatenate the following files keeping variables PatID, CohortInd, and Sequence_Number; creating a new variable, FileInd, and set the value as noted:
    a. PRE1 = Patients_For_NDI_Filled_PHI2 – from package 3
    b. PRE2 = Patients_For_NDI_Research – from Section I in this package, step 2
    c. POST = NDI_Outbound_Records - from Section I in this package, step 4
2. Aggregate this file on FileInd and CohortInd to obtain counts as follows:
    a. Study_Subjects: Number of distinct PatIDs
    b. Alias_Records: Number of records with Sequence_Number greater than "01".
    c. Total_Records: Addition of Study_Subjects and Alias_Records. As a QC step, this value should be equal to the number of records in these source files by FileInd, totaled across both values of CohortInd:
        i. PRE1 = Patients_For_NDI_Filled_PHI2
        ii. PRE2 = Patients_For_NDI_Research
        iii. POST = NDI_Outbound_Records
3. Fill the data dictionary above, sort by FileInd and CohortInd, and save as file ***DPID_SiteID*_Aggregates_Total** to the MSOC data library. DPID_SiteID takes on the value filled in this variable.

IV. Aggregate Descriptive Statistics by Cohort – Outbound to NDI

Using the **NDI_Outbound_Records** file referenced in Section I above NDI Package 1, this aggregate table will be returned to [list lead analytic site] to create a classic "Table 1." Reference the following from NDI Package 1 specifications:

- Section IV Aggregate Descriptive Statistics by Cohort contains the data dictionary and methods for calculations.
- Appendix E displays how this data may be transposed for display purposes and reference.

Methods to Create the Outbound to NDI Aggregate Descriptive Statistics file for [Lead Analytic Site]

4. Link the **NDI_Outbound_Records** from Section I, creating a temporary table as a subset of the **All_Cohorts** file, saved in the DPLocal data library from package 1. The subset for this table is for only those PatIDs from the **All_Cohorts** file that are found in the **NDI_Outbound_Records** file. Note that as the latter file may have more than one row per PatID, the subset created must be distinct by PatID.
5. Using this temporary file, execute the logic from NDI Package 1, Section IV.
6. Name the resulting file as **NDI_Cohort_Descriptives_Outbound** and save to the MSOC data library for this package.

V. Aggregate Statistics of Filled Variables for NDI

This is an aggregate table to indicate the number and percent of records that have NDI identifiers filled and valid.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Full Sentinel DP identifier
FactorID	Num(3)	2.	Factor number	1-15	See Appendix A
NDI_Factor	Char(70)	$70.	Factor/variable assessed	Free text	See Appendix A
Count	Num(4)	comma12.	Count filled	0+	Number of records with variable filled and valid
Percent	Num(4)	percent7.2	Percent filled	0+	Percent of records with variable filled and valid

Methods to create the Aggregate Statistics of Filled Variables for NDI

1. Use the file **NDI_Outbound_Records** saved previously to the DPLocal data library in Section I. Create distinct records by PatID. While this may result in some identifiers varying between rows with the same PatID but with a different Sequence_Number (e.g., Last_Name), the filled identifiers will essentially be the same for purposes of creating this table.
2. Save the total count of resulting rows to a temporary variable AllRows.
3. Fill the variable DPID_SiteID with the value for the health plan.
4. For all factors, perform the following:
    a. Fill FactorID with the value noted in Appendix A.
    b. Fill NDI_Factor with the value from NDI_Factor noted in Appendix A.
5. For each of factors 1-12 perform the following:
    a. Count the number of records with a valid and filled variable as indicated by NDI_Factor; fill the Count variable with this value.
    b. Calculate and fill the Percent variable with Count, from step 5.a, divided by AllRows, as set in step 1.
6. For each of factors 13-15, perform the following:

a. For each record, set a temporary variable to 1, if the record meets the condition specified in Variable_Name and all filled variables in the records are valid, whether part of the condition or not. The condition requires that all variables named must be filled and valid. If the condition is not met, set the temporary variable to 0.
b. Sum the temporary variable across all rows and set Count to this value.
c. Calculate and fill the Percent variable with Count, from step 6.b, divided by AllRows, as set in step 1.

7. Name the resulting file as **NDI_Filled_Factors** and save to the MSOC data library for this package.

Appendix A Filled Variables for NDI

#	NDI_Factor	Variable_Name
1	First name	First_Name
2	Middle initial	Middle_Initial
3	Last name	Last_Name
4	Father's surname	Fathers_Surname
5	Social Security Number (SSN)	SSN
6	Marital status	Marital_Status
7	State of residence	State_of_Residence
8	State of birth	State_of_birth
9	Sex	Sex
10	Month of birth	Month_of_Birth
11	Year of birth	Year_of_Birth
12	Day of birth	Day_of_Birth
13	First AND last name AND month AND year of birth	First_Name AND Last_Name AND Month_of_Birth AND Year_of_Birth
14	First AND last name AND Social Security Number	First_Name AND Last_Name AND SSN
15	Social Security Number, AND month AND day AND year of birth, AND sex	SSN AND Month_of_Birth AND Day_of_Birth AND Year_of_Birth AND Sex

Program Package 5: Read Inbound Files from NDI

I. Quality Control: Inbound Files from NDI – Match, NoMatch, and Rejects

This Program Package processes and checks returned NDI data, and determines the approach for saving the 'best' match from all potential matches.
The NDI returns these three text data files with contents as follows:

- Match: Health plan record matched one or more NDI records
- NoMatch: Health plan record did not match any NDI records
- Rejects: Health plan record was rejected prior to attempting a match to NDI records

Each of these three files is in the exact same format as the submitted text data file. See data dictionary specifications for Package 4, Section I.

This section of the program package will assess the returned files against the submitted file and fill the data dictionary as shown. The intent is to ensure that all outbound records from health plans to the NDI are found in one of the three returned files from the NDI and that no additional records have been returned.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
Control_Number	Char(10)	$10.	Control number	0000000001+	Sequential number, with leading zeros, incremented for each PatID written to this table, inserted automatically
Sequence_Number	Char(2)	$2.	Sequence number	01+	Sequential number, with leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
NDI_Source	Char(2)	$2	NDI source file code	M= Match file N= NoMatch file R= Reject file NF= Not found in any NDI file	Code for source of NDI file
Link	Char(1)	$1.	Link code	B= Found in both NDI and health plan files H= Found only in health plan file N= Found only in NDI file	Quality control indicator of links between NDI returned files and health plan source files

Methods to create the patient-level NDI quality control file

1. Check that the file named **NDI_Outbound_Records** is found in the DPLocal data library from Package 4.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal library.
    b. Otherwise if found, proceed with all other processes.
2. One or two *sets* of three text files are returned by the NDI to health plans: Match, NoMatch, and Rejects. One set is always returned for "routine" searches (up through 2017 death data); some health plans will have a second set returned for "early release" (2018 death data) searches. The extension ".TXT" is not part of the filename. Establish parameters for the name of each of the three files, for each of the two sets, and set the parameters to the names indicated. Health plans will be instructed to place these files in the DPLocal data library for this package 5 and to check that the parameters match the names of the actual files returned.
    a. The following plans are expected to have only one set of files: KPCO and VB.
    b. The following plans are expected to have only two sets of files: Aetna, KPWA/GHC, HP, and KPNW.
3. The following are examples of expected filenames:

   a. Routine: 2018-X010F00.MATCH, 2018-X010F00.NOMATCH, and 2018-X010F00.REJECTS
   b. Early release: ER2018-X010F00.MATCH, ER2018-X010F00.NOMATCH, and ER2018-X010F00.REJECTS
4. Check that each of these 3 (or 6) files are found in the DPLocal data library for this programming Package 5.
   a. If any are not found, abort this process and print an error message to the log indicating which of the 3 (or 6) file(s) are missing from the DPLocal data library.
   b. Otherwise if found, proceed with all other processes.
5. Use the NDI Match, NoMatch, and Rejects text files. If the health plan has 2 sets, concatenate the results across each of the 2 sets into the 3 files: Match, NoMatch, and Rejects. From each of the three files, read in only the following column positions to create variables as shown:
   a. Control_Number: Column positions 82-91
   b. Sequence_Number: Column positions 92-93
6. In each file, set a variable NDI_Source, setting the single character values shown in the data dictionary above.
7. Concatenate the three files into a temporary file.
8. Using variables Control_Number and Sequence_Number, perform a full merge (i.e., SQL full join) of the temporary file in step 7 to the **NDI_Outbound_Records** file, found in the DPLocal data library for Package 4. Add variable PatID to the file. Note that PatID may appear on more than one row in the **NDI_Outbound_Records** file, but rows should be unique on the basis of PatID/Sequence_Number.
9. Set variable Link as follows:
   a. B= The row was found in both files.
   b. N= The row was found only in the NDI file.
   c. H= The row was found only in the health plan file.
10. If Link=H but there is no corresponding NDI record, then set NDI_Source=NF.
11. Save the file as **NDI_Quality_Control** to the DPLocal data library.

II. Inbound File from NDI - Combined

The following is the data dictionary of patient level data to be imported as flat text files, transformed into SAS datasets, from the NDI Combined file. The Combined file contains information only for matches to NDI records. This file combines the information provided on a user record with information on the matching NDI record(s). Each Combined record on this file contains information from only one user record and only one matching NDI record. The first 100 positions of each Combined record contain the actual data from the user record. Variables with filled values for the "Column Positions" column are to be imported as character or numeric, as specified.

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	N/A, based on link	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
Last_Name	1-20	Char(20)	$20.	Last name of patient	Alpha chars, blank, <=1 of ' and/or -)	Left justified: Last Name of Patient to be provided by the Data Partner.
First_Name	21-35	Char(15)	$15.	First name of patient	Alpha chars, blank	Left justified: First Name of Patient to be provided by Data Partner.
Middle_Initial	36	Char(1)	$1.	Middle initial of patient name	Alpha chars, blank	Patient's middle initial to be provided by Data Partner.
SSN	37-45	Char(9)	$9.	Social Security Number	9 numeric digits, blank	Social Security Number, exactly 9 digits, maintaining leading zeros, to be provided by Data Partner.
Month_of_Birth	46-47	Num(3)	Z2.	Month of birth	01-12	Patient's month of birth, with leading zero, automatically filled from SDD.
Day_of_Birth	48-49	Num(3)	Z2.	Day of birth	01-31	Patient's day of birth, with leading zero, automatically filled from SDD.

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Year_of_Birth	50-53	Num(3)	4.	Year of birth	1850-20##	Patient's year of birth, automatically filled from SDD.
Fathers_Surname	54-71	Char(18)	$18.	Fathers surname	Alpha chars, <= 1 of ' and/or -, or blank	Left justified: Patient's father's surname to be provided by Data Partner.
Age_Unit	72	Num(3)	1.	Age unit recorded (at death)	0= Died < 100 years 1=Died>= 100 years or Blank	Unit the age is recorded in, unknown and will not be populated (left blank).
Number_Age_Units	73-74	Num(3)	Z2.	Number of age units (at death)	0-99, blank	Number of age units at time of death; unknown will not be populated (left blank).
Sex	75	Char(1)	$1.	Sex of patient	M, F	Sex of patient, automatically recoded and filled from SDD.
Race	76	Char(1)	$1.	Race of patient	0-8, blank	Race of patient, automatically recoded and filled from SDD.
Marital_Status	77	Char(1)	$1.	Marital status of patient	1-4, blank	Patient's marital status to be provided by Data Partner
State_of_Residence	78-79	Char(2)	$2.	State of residence	See Appendix A	Patient's state of residence automatically recoded and filled with last known value in SDD[1].
State_of_birth	80-81	Char(2)	$2.	State of birth	See Appendix A	State where patient was born, unknown and will not be populated (left blank).
Control_Number	82-91	Char(10)	$10.	Control number	0000000001+	Sequential number, with leading zeros, incremented for each PatID written to this table, inserted automatically

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Sequence_Number	92-93	Char(2)	$2.	Sequence number	01+	Sequential number, with leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"
Blank	94-100					These column positions are blank; no variable to be imported
State_Of_Death	101-112	Char(12)	$12.	State of death	Text state names	NDI
Year_Of_Death	113-116	Num(3)	4.	4-digit year	2000-2017	
State_Death_Code	117-119	Char(3)	$3.	State code	See Appendix A	01–57, 59, left justified; 33C if New York City. Numeric codes are converted to alpha codes
Alias	120	Char(1)	$1.	Alias	A or blank	"A" if an alias record; otherwise, blank
Death_Cert	121-126	Char(6)	$6.	Death certificate number	Free text	State death certificate number
NDI_Death_Date_MO	127-128	Num(3)	2.	Death date month	1-12	
NDI_Death_Date_DAY	129-130	Num(3)	2.	Death date day	1-31	
NDI_Death_Date_YR	131-132	Num(3)	z2.	Death date year	00-99	
NDI_Death_Date	127-132	Num(4)	mmddyy10.	Death date	Valid SAS dates	Constructed from date parts
Match_First	133-134	Char(2)	$2.	Match on first name	bX, bN, b?, lb, IN, bb (b = blank space)	Indicates quality of match on first name

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Match_Middle	135	Char(1)	$1.	Match on middle initial	X, N, B, -, ?, or blank	Indicates quality of match on middle initial
Match_Last	136	Char(1)	$1.	Match on last name	X, N, ?, or blank	Indicates quality of match on last name
Match_Father_Surname	137	Char(1)	$1.	Match on father surname	X, N, ?, -, or blank	Indicates quality of match on father surname
Match_Last_Surname	138	Char(1)	$1.	Match on last to father name	X, N, ?, -, or blank	Indicates quality of match on last name to father name on NDI record
Match_SSN	139-147	Char(9)	$9.	Match on SSN	X or - in each position, or single – in position 143, or single ? in position 143	Indicates quality of match on SSN
Match_Birth_Month	148	Char(1)	$1.	Match on birth month	X, ?, -, or blank	Indicates quality of match on birth month
Match_Birth_Day	149	Char(1)	$1.	Match on birth day	X, ?, -, or blank	Indicates quality of match on birth day
Match_Birth_Year	150-152	Char(3)	$3.	Match on birth year	Xbb, ?bb, -b (b = blank space), or +01, -01, +02, -02, … +99, -99, >99	Indicates quality of match on birth year
Match_Age	153	Char(1)	$1.	Match on age at death	X, ?, -, blank	Indicates quality of match on age at death
Match_Sex	154	Char(1)	$1.	Match on sex	X, ?, -, blank	Indicates quality of match on sex
Match_Race	155	Char(1)	$1.	Match on race	X, ?, -, blank	Indicates quality of match on race

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
Match_Marital	156	Char(1)	$1.	Match on marital status	X, ?, -, blank	Indicates quality of match on marital status
Match_State_Residence	157	Char(1)	$1.	Match on residence state	X, ?, -, blank	Indicates quality of match on state of residence
Match_State_Birth	158	Char(1)	$1.	Match on birth state	X, ?, -, blank	Indicates quality of match on state of birth
Blank	159-164					*These column positions are blank; no variable to be imported*
Exact	165	Char(1)	$1.	Exact match indicator	* (asterisk) or blank	Indicates exact match of data to NDI record
Matching_Sequence	166-168	Num(3)	2.	NDI matching record sequence	Read in numerics with leading zeros from 1-50	NDI matching sequence indicator
Possible_Sequence	169-171	Num(3)	2.	NDI possible sequences	Read in numerics with leading zeros from 1-50	NDI maximum number of records indicator
Prob_Score	172-176	Num(4)	5.2	Probabilistic matching score	##.## format	NDI matching score
Class_Code	177	Num(3)	1.	Class code	1-5	NDI class code
Status_Code	178	Num(3)	1.	Status code	0=False (presumed alive) 1=True (presumed dead)	

Methods to create the patient-level Combined file after receipt from the NDI

1. Check that the text file named **Combined** (e.g., 2018-X010F00.COMBINED for all plans; ER2018-X010F00.COMBINED for plans identified with early release; see Section I) is/are found in the DPLocal data library for this programming Package 5.
    a. If one or both files, as expected are not found, abort this process and print an error message to the log indicating which file(s) must be in the DPLocal data library.
    b. Otherwise if found, proceed with all other processes.
2. If two files are found, concatenate into a single Combined file. Use the NDI Combined text file. Read into a SAS dataset that meets the requirements of the data dictionary.
3. Link this file, by Control_Number and Sequence_Number, to the **NDI_Outbound_Records** file, found in the DPLocal data library from Package 4. Add variable PatID to the file. Note that PatID may appear on more than one row, but rows should be unique on the basis of PatID/Sequence_Number.
4. For State_Death_Code, using the numeric postal code in the NDI inbound source data with (See Appendix A) convert to alpha codes.
    a. Set “33C” to “NYC”.
    b. Set “99” to “UN”.
5. Use the variables NDI_Death_Date_MO, NDI_Death_Date_DAY, and NDI_Death_Date_YR to create variable NDI_Death_Date. If invalid dates get created, then set NDI_Death_Date to missing.
6. Check concordance with Match file: Create a temporary file of the current Combined file, resolving to rows with distinct values of only Control_Number and Sequence_Number. Filter the **NDI_Quality_Control** file, from Section I, for records where NDI_Source=M, keeping variables Control_Number and Sequence_Number.
    a. Perform a full merge of the temporary Combined file and the filtered NDI_Quality_Control file on Control_Number and Sequence_Number. The merge should result in a 1:1 relationship between rows from both files.
    b. If there is not a 1:1 relationship between the two files, perform the following steps. Otherwise, go to step 7.
        i. Create a file of the full merge, saved to DPLocal and called **MisMatch_Combined_Match**, containing the following variables: PatID, Control_Number, and Sequence_Number. Create an additional variable called Status (character of length 1) that will contain the following values:
            1. B=Found in both files
            2. C=Found in only the Combined file
            3. M=Found in only the Match file
        ii. Write a message to the log that there is a mismatch between records in the Match and Combined files.

iii. Create a temporary switch variable called Mismatch_Abort that can contain values of Y or No. If the switch is set to Y, then abort the program.

7. Name the file as **NDI_Combined** and save the file to the DPLocal data library established for this project.

III. Inbound File from NDI - Cause

The following is the data dictionary of patient level data to be imported from a flat text file, transformed into SAS datasets, from the NDI Cause file. The Cause file contains information only for matches to NDI records which are rated as True matches, ranked first, or had a high probabilistic score. The Cause file begins with the same 178-record format as the Combined file (positions 1–178). The remaining fields on the Cause file contain the coded causes of death (see positions 180– 438). The order of the user records in the Cause file will be the same as the order in the Combined file.

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Based on link	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
- All other variables in the NDI_Combined file (Positions 1-178) -						
Blank	179					This column position is blank; no variable to be imported
Cause_Underlying	180-183	Char(4)	$4.	Underlying cause of death	ICD10	
Cause_Recode_358	184-188	Char(5)	$5.	Recoded underlying cause of death	ICD10	Recode into 358 ICD10 groups
Cause_Recode_113	189-191	Char(3)	$3.	Recoded underlying cause of death	ICD10	Recode into 113 ICD10 groups
Cause_Recode_130	192-194	Char(3)	$3.	Recoded underlying cause of death	ICD10	Recode into 130 ICD10 groups
Number_Axis	195-196	Num(3)	2.	Axis number	0-20	Number of entity axis conditions; recode from text values with leading zeros
EntityPosition1	197	Char(1)	$1.	Entity position 1	1-6	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntitySequence1	198	Num(3)	1.	Entity sequence 1	1-7	
EntityCode1	199-202	Char(4)	$4.	Entity code 1	ICD10 codes	
EntityInjury1	203	Char(1)	$1.	Entity injury flag 1	1=Nature of injury 0=All other codes	
EntityPosition2	204	Char(1)	$1.	Entity position 2	1-6	
EntitySequence2	205	Num(3)	1.	Entity sequence 2	1-7	
EntityCode2	206-209	Char(4)	$4.	Entity code 2	ICD10 codes	
EntityInjury2	210	Char(1)	$1.	Entity injury flag 2	1=Nature of injury 0=All other codes	
EntityPosition3	211	Char(1)	$1.	Entity position 3	1-6	
EntitySequence3	212	Num(3)	1.	Entity sequence 3	1-7	
EntityCode3	213-216	Char(4)	$4.	Entity code 3	ICD10 codes	
EntityInjury3	217	Char(1)	$1.	Entity injury flag 3	1=Nature of injury 0=All other codes	
EntityPosition4	218	Char(1)	$1.	Entity position 4	1-6	
EntitySequence4	219	Num(3)	1.	Entity sequence 4	1-7	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityCode4	220-223	Char(4)	$4.	Entity code 4	ICD10 codes	
EntityInjury4	224	Char(1)	$1.	Entity injury flag 4	1=Nature of injury 0=All other codes	
EntityPosition5	225	Char(1)	$1.	Entity position 5	1-6	
EntitySequence5	226	Num(3)	1.	Entity sequence 5	1-7	
EntityCode5	227-230	Char(4)	$4.	Entity code 5	ICD10 codes	
EntityInjury5	231	Char(1)	$1.	Entity injury flag 5	1=Nature of injury 0=All other codes	
EntityPosition6	232	Char(1)	$1.	Entity position 6	1-6	
EntitySequence6	233	Num(3)	1.	Entity sequence 6	1-7	
EntityCode6	234-237	Char(4)	$4.	Entity code 6	ICD10 codes	
EntityInjury6	238	Char(1)	$1.	Entity injury flag 6	1=Nature of injury 0=All other codes	
EntityPosition7	239	Char(1)	$1.	Entity position 7	1-6	
EntitySequence7	240	Num(3)	1.	Entity sequence 7	1-7	
EntityCode7	241-244	Char(4)	$4.	Entity code 7	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityInjury7	245	Char(1)	$1.	Entity injury flag 7	1=Nature of injury 0=All other codes	
EntityPosition8	246	Char(1)	$1.	Entity position 8	1-6	
EntitySequence8	247	Num(3)	1.	Entity sequence 8	1-7	
EntityCode8	248-251	Char(4)	$4.	Entity code 8	ICD10 codes	
EntityInjury8	252	Char(1)	$1.	Entity injury flag 8	1=Nature of injury 0=All other codes	
EntityPosition9	253	Char(1)	$1.	Entity position 9	1-6	
EntitySequence9	254	Num(3)	1.	Entity sequence 9	1-7	
EntityCode9	255-258	Char(4)	$4.	Entity code 9	ICD10 codes	
EntityInjury9	259	Char(1)	$1.	Entity injury flag 9	1=Nature of injury 0=All other codes	
EntityPosition10	260	Char(1)	$1.	Entity position 10	1-6	
EntitySequence10	261	Num(3)	1.	Entity sequence 10	1-7	
EntityCode10	262-265	Char(4)	$4.	Entity code 10	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityInjury10	266	Char(1)	$1.	Entity injury flag 10	1=Nature of injury 0=All other codes	
EntityPosition11	267	Char(1)	$1.	Entity position 11	1-6	
EntitySequence11	268	Num(3)	1.	Entity sequence 11	1-7	
EntityCode11	269-272	Char(4)	$4.	Entity code 11	ICD10 codes	
EntityInjury11	273	Char(1)	$1.	Entity injury flag 11	1=Nature of injury 0=All other codes	
EntityPosition12	274	Char(1)	$1.	Entity position 12	1-6	
EntitySequence12	275	Num(3)	1.	Entity sequence 12	1-7	
EntityCode12	276-279	Char(4)	$4.	Entity code 12	ICD10 codes	
EntityInjury12	280	Char(1)	$1.	Entity injury flag 12	1=Nature of injury 0=All other codes	
EntityPosition13	281	Char(1)	$1.	Entity position 13	1-6	
EntitySequence13	282	Num(3)	1.	Entity sequence 13	1-7	
EntityCode13	283-286	Char(4)	$4.	Entity code 13	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityInjury13	287	Char(1)	$1.	Entity injury flag 13	1=Nature of injury 0=All other codes	
EntityPosition14	288	Char(1)	$1.	Entity position 14	1-6	
EntitySequence14	289	Num(3)	1.	Entity sequence 14	1-7	
EntityCode14	290-293	Char(4)	$4.	Entity code 14	ICD10 codes	
EntityInjury14	294	Char(1)	$1.	Entity injury flag 14	1=Nature of injury 0=All other codes	
EntityPosition15	295	Char(1)	$1.	Entity position 15	1-6	
EntitySequence15	296	Num(3)	1.	Entity sequence 15	1-7	
EntityCode15	297-300	Char(4)	$4.	Entity code 15	ICD10 codes	
EntityInjury15	301	Char(1)	$1.	Entity injury flag 15	1=Nature of injury 0=All other codes	
EntityPosition16	302	Char(1)	$1.	Entity position 16	1-6	
EntitySequence16	303	Num(3)	1.	Entity sequence 16	1-7	
EntityCode16	304-307	Char(4)	$4.	Entity code 16	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityInjury16	308	Char(1)	$1.	Entity injury flag 16	1=Nature of injury 0=All other codes	
EntityPosition17	309	Char(1)	$1.	Entity position 17	1-6	
EntitySequence17	310	Num(3)	1.	Entity sequence 17	1-7	
EntityCode17	311-314	Char(4)	$4.	Entity code 17	ICD10 codes	
EntityInjury17	315	Char(1)	$1.	Entity injury flag 17	1=Nature of injury 0=All other codes	
EntityPosition18	316	Char(1)	$1.	Entity position 18	1-6	
EntitySequence18	317	Num(3)	1.	Entity sequence 18	1-7	
EntityCode18	318-321	Char(4)	$4.	Entity code 18	ICD10 codes	
EntityInjury18	322	Char(1)	$1.	Entity injury flag 18	1=Nature of injury 0=All other codes	
EntityPosition19	323	Char(1)	$1.	Entity position 19	1-6	
EntitySequence19	324	Num(3)	1.	Entity sequence 19	1-7	
EntityCode19	325-328	Char(4)	$4.	Entity code 19	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
EntityInjury19	329	Char(1)	$1.	Entity injury flag 19	1=Nature of injury 0=All other codes	
EntityPosition20	330	Char(1)	$1.	Entity Position 20	1-6	
EntitySequence20	331	Num(3)	1.	Entity Sequence 20	1-7	
EntityCode20	332-335	Char(4)	$4.	Entity Code 20	ICD10 codes	
EntityInjury20	336	Char(1)	$1.	Entity Injury Flag 20	1=Nature of injury 0=All other codes	
Number_Record	337-338	Num(3)	2.	# of Record Axis Conditions	0-20	
RecordCode1	339-342	Char(4)	$4.	Record code 1	ICD10 codes	
RecordInjury1	343	Char(1)	$1.	Record injury flag 1	1=Nature of injury 0=All other codes	
RecordCode2	344-347	Char(4)	$4.	Record code 2	ICD10 codes	
RecordInjury2	348	Char(1)	$1.	Record injury flag 2	1=Nature of injury 0=All other codes	
RecordCode3	349-352	Char(4)	$4.	Record code 3	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
RecordInjury3	353	Char(1)	$1.	Record injury flag 3	1=Nature of injury 0=All other codes	
RecordCode4	354-357	Char(4)	$4.	Record code 4	ICD10 codes	
RecordInjury4	358	Char(1)	$1.	Record injury flag 4	1=Nature of injury 0=All other codes	
RecordCode5	359-362	Char(4)	$4.	Record code 5	ICD10 codes	
RecordInjury5	363	Char(1)	$1.	Record injury flag 5	1=Nature of injury 0=All other codes	
RecordCode6	364-367	Char(4)	$4.	Record code 6	ICD10 codes	
RecordInjury6	368	Char(1)	$1.	Record injury flag 6	1=Nature of injury 0=All other codes	
RecordCode7	369-372	Char(4)	$4.	Record code 7	ICD10 codes	
RecordInjury7	373	Char(1)	$1.	Record injury flag 7	1=Nature of injury 0=All other codes	
RecordCode8	374-377	Char(4)	$4.	Record code 8	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
RecordInjury8	378	Char(1)	$1.	Record injury flag 8	1=Nature of injury 0=All other codes	
RecordCode9	379-382	Char(4)	$4.	Record code 9	ICD10 codes	
RecordInjury9	383	Char(1)	$1.	Record injury flag 9	1=Nature of injury 0=All other codes	
RecordCode10	384-387	Char(4)	$4.	Record code 10	ICD10 codes	
RecordInjury10	388	Char(1)	$1.	Record injury flag 10	1=Nature of injury 0=All other codes	
RecordCode11	389-392	Char(4)	$4.	Record code 11	ICD10 codes	
RecordInjury11	393	Char(1)	$1.	Record injury flag 11	1=Nature of injury 0=All other codes	
RecordCode12	394-397	Char(4)	$4.	Record code 12	ICD10 codes	
RecordInjury12	398	Char(1)	$1.	Record injury flag 12	1=Nature of injury 0=All other codes	
RecordCode13	399-402	Char(4)	$4.	Record code 13	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
RecordInjury13	403	Char(1)	$1.	Record injury flag 13	1=Nature of injury 0=All other codes	
RecordCode14	404-407	Char(4)	$4.	Record code 14	ICD10 codes	
RecordInjury14	408	Char(1)	$1.	Record injury flag 14	1=Nature of injury 0=All other codes	
RecordCode15	409-412	Char(4)	$4.	Record code 15	ICD10 codes	
RecordInjury15	413	Char(1)	$1.	Record injury flag 15	1=Nature of injury 0=All other codes	
RecordCode16	414-417	Char(4)	$4.	Record code 16	ICD10 codes	
RecordInjury16	418	Char(1)	$1.	Record injury flag 16	1=Nature of injury 0=All other codes	
RecordCode17	419-422	Char(4)	$4.	Record code 17	ICD10 codes	
RecordInjury17	423	Char(1)	$1.	Record injury flag 17	1=Nature of injury 0=All other codes	
RecordCode18	424-427	Char(4)	$4.	Record code 18	ICD10 codes	

Variable	Column Positions	Type (Length)	Format	Label	Valid Values	Source / Comments
RecordInjury18	428	Char(1)	$1.	Record injury flag 18	1=Nature of injury 0=All other codes	
RecordCode19	429-432	Char(4)	$4.	Record code 19	ICD10 codes	
RecordInjury19	433	Char(1)	$1.	Record injury flag 19	1=Nature of injury 0=All other codes	
RecordCode20	434-437	Char(4)	$4.	Record code 20	ICD10 codes	
RecordInjury20	438	Char(1)	$1.	Record injury flag 20	1=Nature of injury 0=All other codes	

Methods to create the patient-level Cause file after receipt from the NDI

1. Check that the text file named **Cause** (e.g., 2018-X010F00.CAUSE for all plans; ER2018-X010F00.CAUSE for plans identified with early release; see Section I) is/are found in the DPLocal data library for this programming Package 5.
    a. If one or both files, as expected are not found, abort this process and print an error message to the log indicating which file(s) must be in the DPLocal data library.
    b. Otherwise if found, proceed with all other processes.
2. If two files are found, concatenate into a single Cause file. Use the NDI Cause text file. Read into a SAS dataset that meets the requirements of the data dictionary.
3. Link this file, by Control_Number and Sequence_Number, to the **NDI_Outbound_Records** file, found in the DPLocal data library for Package 4. Add variable PatID to the file. Note that PatID may appear on more than one row.

4. For State_code, using the numeric postal code in the NDI inbound source data with Appendix A, convert to alpha codes.
    a. Set “33C” to “NYC”.
    b. Set “99” to “UN”.
5. Use the variables NDI_Death_Date_MO, NDI_Death_Date_DAY, and NDI_Death_Date_YR to create variable NDI_Death_Date. If invalid dates get created, then set NDI_Death_Date to missing.
6. Name the file as **NDI_Cause** and save the file to the DPLocal data library established for this project.

IV. Selection of NDI Records

It is possible that for any patient record sent to the NDI, there can be no matches, a match with one NDI record, or a match with multiple NDI records (e.g., one cohort member (i.e., PatID) matches multiple death certificates in NDI data). This process will identify which NDI matched record shall be kept for those patient records that had a match to a NDI record. For some cohort members, there will be ties to multiple NDI records and these will be saved in a separate file.

This process will also provide information to the study team regarding the quality of matches (e.g., distribution of Status code, Class, and Probabilistic score). Selection shall be based on each of three different algorithms. Multiple patient-level files are created by the processes in this step and all follow the same data dictionary:

- **NDI_Selection:** One row per PatID and death certificate number
- **Potential_Ties:** Two or more rows per PatID
- **Duplicates:** One or more rows per PatID

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char (Site specific length)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
Control_Number	Char(10)	$10.	Control number	0000000001+	Sequential number, leading zeros, incremented for each PatID written to this table, inserted automatically
Sequence_Number	Char(2)	$2.	Sequence number	01+	Sequential number, leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Year_NDI	Num(3)	4.	NDI year to search	2000-2017	NDI death year to be searched, based on DeathDt or DisenrollDate
Reason_Select	Char(3)	$3.	Reason patient was sent to NDI	DE=Death DI=Disenrollment B=Death and disenrollment LTF=Lost to follow-up	
COD_Available	Num(3)	1.	COD available	1=Available 0=Not available	Indicates whether patient row has a row in the NDI Cause table
Death_Date_Complete	Num(3)	1.	Full death date filled	1=Yes, filled 0=No, not filled	Indicator if a full death was supplied by NDI
Death_Month_Complete	Num(3)	1.	Death month filled	1=Yes, filled 0=No, not filled	Indicator if death month was supplied by NDI
Death_Day_Complete	Num(3)	1.	Death day filled	1=Yes, filled 0=No, not filled	Indicator if death day was supplied by NDI
Death_Year_Complete	Num(3)	1.	Death year filled	1=Yes, filled 0=No, not filled	Indicator if a death year was supplied by NDI
NDI_Death_Date	Num(4)	mmddyy10.	Death date	Valid SAS dates	
Death_Cert	Char(6)	$6.	Death certificate number	Free text	State death certificate number

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
State_Death_Code	Char(3)	$3.	State code	See Appendix A	01–57, 59, left justified; 33C if New York City. Numeric codes are converted to alpha codes
Algorithm1	Num(3)	1.	Algorithm 1	1=Selected 0=Not selected	Indicates whether patient row outbound to NDI has a selected match, based on algorithm 1
Algorithm2	Num(3)	1.	Algorithm 2	1=Selected 0=Not selected	Indicates whether patient row outbound to NDI has a selected match, based on algorithm 2
Algorithm3	Num(3)	1.	Algorithm 3	1=Selected 0=Not selected	Indicates whether patient row outbound to NDI has a selected match, based on algorithm 3
Exact	Char(1)	$1.	Exact match indicator	* (asterisk) or blank	Indicates exact match of all health plan data to NDI record
Matching_Sequence	Num(3)	2.	NDI matching record sequence	Read in numerics with leading zero from 1-50	NDI matching sequence indicator
Prob_Score	Num(4)	5.2	Probabilistic matching score	##.## format	NDI matching score
Class_Code	Num(3)	1.	Class code	1-5	NDI class code
Status_Code	Num(3)	1.	Status code	0=False (presumed alive) 1=True (presumed dead)	NDI status code

Methods for selection of NDI records

1. Check that the file named **All_Cohorts** is found in the DPLocal data library for Package 1.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 1.
    b. Otherwise if found, proceed with all other processes.
2. Check that the file named **R01_Signature** is found in the MSOC data library from Package 1.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the MSOC data library.
    b. Otherwise if found, proceed as follows:
        i. Scan for the row where Var = "DPMaxDate".
        ii. Create a temporary variable (e.g., macro variable), DPMaxDate, and set it to the date value of the character value found in the Value variable. This will be used in steps below.
3. Use the **Combined** file created above, with variables PatID, Control_Number, Sequence_Number,COD_Available, and NDI_Death_Date, Death_Cert, and State_Death_Code, as source for all of the following steps.
4. Determine if the health plan identifies a file that lists PatIDs not eligible for research.
    a. If so, then on PatID, merge the **Combined** file with such file, keeping all rows from **Combined** containing PatIDs not found in the health plan file that lists PatIDs not eligible for research. Save this file temporarily as **Combined_Temp**; continue with step 5.
    b. Otherwise, copy **Combined** to temporary file **Combined_Temp** and continue with next steps.
5. Then link, on Control_Number and Sequence_Number with the **Cause** table, keeping all rows from the **Combined_Temp** file.
    a. If there is a corresponding row in the Cause table, then set COD_Available to 1.
    b. Otherwise, set COD_Available to 0.
    c. Keep all variables from the Combined file as well as only the COD_Available variable.
6. Link on PatID with the **All_Cohorts** file from Package 1 to obtain variables IndexDt, CohortInd, DeathDt, DisenrollDate, DeathSrc, and FinalEnrollmentStatus for the next steps.
7. Create variable Reason_Select, using variables DeathSrc and FinalEnrollmentStatus as follows:
    a. DE = DeathSrc equal to D, E, or B
    b. DI = FinalEnrollmentStatus equal to D
    c. B = Both steps 7a and 7b are true
    d. LTF (for "lost to follow-up" = Neither steps 7a nor 7b are true)
8. Set Year_NDI as follows:

   a. If DeathDt is filled, then fill as year portion of DeathDt.
   b. If DeathDt is not filled, then fill with year portion of DisenrollDate.
   c. If both DeathDt and DisenrollDate are not filled, then set to the year value of temporary variable DPMaxDate.

9. Set Death_Date_Complete as follows:
   a. 1=NDI_Death_Date is a complete date.
   b. 0=NDI_Death_Date is missing.
10. Set Death_Month_Complete as follows:
   a. 1=NDI_Death_Date_MO contains values 1-12.
   b. 0=NDI_Death_Date_MO contains values other than 1-12.
11. Set Death_Day_Complete as follows:
   a. 1=NDI_Death_Date_DAY contains values 1-31.
   b. 0= NDI_Death_Date_DAY contains values other than 1-31.
12. Set Death_Year_Complete as follows:
   a. 1=NDI_Death_Date_YR contains values 00-99.
   b. 0= NDI_Death_Date_YR contains values other than 00-99.
13. Set each of variables Algorithm1, Algorithm2, and Algorithm3 to 0, for records that meet any of the following conditions:
   a. NDI_Death_Date is earlier than IndexDt
   b. Prob_Score < 0
   c. Class_Code = 5

   Saved these records to the **NDI_Selection** file.

14. All other records will have one of the following 3 outcomes:
   a. Duplicates: These are defined as multiple cases of same PatID/Death_Cert combinations, keeping from those records, the record with the Exact indicator, followed by that with the lowest Class_Code and highest Prob_Score value. Set Algorithm1=0 for the other record(s) and save to file **Duplicates**.
      i. If the same PatID/Death_Cert has the *same* value for the Exact indicator, Class_Code and Prob_Score, then select the record to keep at random for *further processing below*. Save all selected records at this stage to **Combined_Distinct**; this file will be used in Section VII below. Set Algorithm1=0 for the other record and save to file **Duplicates**.
   b. Tied records saved to **Potential_Ties**: Further defined below, these are records with the same values for Exact indicator, Class_Code, and Prob_Score. They will be saved to (or updated within) file **Potential_Ties** depending on logic in Algorithms 1-3.

c. Distinct records saved to **NDI_Selection** file: Selected records with Algorithm1=1 are saved to (or updated within) this file. Non-selected records with Algorithm1=0 (including dropped duplicates as per step 14.a) are also saved to this file.

15. Proceed with the following steps for all distinct records, defined as those with same PatID but differing Death_Cert.
16. **Algorithm 1:** This is an NDI recommended method for selecting matched records.
    a. Select records where Status_Code=1 (True match; assumed dead) and Class_Code is between 1 and 4 inclusive.
    b. For those with Class_Code = 1:
        i. If for any PatID, there is more than 1 NDI record with Class_Code=1, then retain the record with the highest value of Prob_Score.
        ii. Then if for any PatID, there is more than 1 NDI record with Class_Code=1 and there are identical values of Prob_Score, then select the record with Exact="*".
        iii. Then if for any PatID, there is more than 1 NDI record, with Class_Code=1, identical values of Prob_Score, and all have Exact="*", set Algorithm1=1 to indicate the algorithm being evaluated, and write these records to file **Potential_Ties**.
    c. For those with Class_Code = 2:
        i. Keep only those records with Prob_Score >= 44.5.
        ii. If for any PatID, there is more than 1 NDI record with Class_Code=2 and Prob_Score >= 44.5, then keep the record with the highest value of Prob_Score.
        iii. If for any PatID, there is more than 1 NDI record with Class_Code=2 and Prob_Score >= 44.5, set Algorithm1=1 to indicate the algorithm being evaluated, and write these records to file **Potential_Ties**.
    d. For those with Class_Code = 3:
        i. Keep only those records with Prob_Score >= 37.5.
        ii. If for any PatID, there is more than 1 NDI record with Class_Code=3 and Prob_Score >= 37.5, then keep the record with the highest value of Prob_Score.
        iii. If for any PatID, there is more than 1 NDI record with Class_Code=3 and Prob_Score >= 37.5, set Algorithm1=1 to indicate the algorithm being evaluated, and write these records to file **Potential_Ties**.
    e. For those with Class_Code = 4:
        i. Keep only those records with Prob_Score >= 32.5.
        ii. If for any PatID, there is more than 1 NDI record with Class_Code=4 and Prob_Score >= 32.5, then keep the record with the highest value of Prob_Score.
        iii. If for any PatID, there is more than 1 NDI record with Class_Code=4 and Prob_Score >= 32.5, set Algorithm1=1 to indicate the algorithm being evaluated, and write these records to file **Potential_Ties**.

f. For each PatID that may exist in multiple steps 16.b through 16.e, save the record with the lowest Class_Code and highest Prob_Score. For all selected records within this step 14, set Algorithm1 equal to 1. Set Algorithm1 equal to 0 for all other records. Save records to the **NDI_Selection** table.

17. **Algorithm 2:**
    a. Keep only records returned with a Status_Code = 1 or 0. Of these, then keep matches with a Class_Code between 1 and 4 inclusive.
    b. Follow steps 16.b through 16.f above, except set Algorithm2 to 1 or 0 as required, following step 16.f, for records written to **NDI_Selection** and tied records written to **Potential_Ties**.
18. **Algorithm 3:** Creates a separate score variable (variable Project_Score), assigns the lowest Project_Score variable to each record, and selects the match on the basis of the Project_Score and the NDI's Prob_Score variables.
    a. SSN management: Within Match_SSN, count the number of "X" values and assign count to SSN_Score.
        i. If SSN:
            1. Has any of these values: 000999999, 009999999, 078051120, 099999999, 123456789, 888888888, 999999990, 999999999
            2. Is in specific range of: 987-65-4320 through 987-65-4329
            3. Contains invalid values in positions 1-3 (area numbers): 000, 666 and any of 900-999
            4. Contains invalid values in positions 4-5 (group numbers): 00
            5. Contains invalid values in positions 6-9 (serial numbers): 0000

            Then set SSN_Score to missing.

        ii. Check SSN for typos: If SSN_Score=7 and Match_SSN contains 2 hyphens in succession, set Typo=1; otherwise, set Typo=0.
        iii. Create flag for "close" SSN: If SSN_Score >= 8 or Typo=1, set SSN_Close=1; otherwise set SSN_Close=0.
    b. Birthdate management: Perform the next steps in order.
        i. Full birth date agreement: If Match_Birth_Month, Match_Birth_Day, and Match_Birth_Year all = X, then set DOB_Agree=1.
        ii. Birth date month and year agreement: If Match_Birth_Month and Match_Birth_Year both = X, then set DOB_MoYr=1.
        iii. Birth year agrees within 3 years; month and day have complete agreement: If Match_Birth_Month and Match_Birth_Day both = X, and Match_Birth_Year contains any of -03, -02, -01, +01, +02, or +03, then set DOB_3Year=1.
        iv. Two out of three parts of birth date match: If any *pairs* of these variables (Match_Birth_Month, Match_Birth_Day, or Match_Birth_Year) both contain X, then set DOB_2of3 =1.
    c. Name management: Perform the next steps in order.
        i. First and last names match: If Match_First and Match_Last both = X, then set Name_FullAgree/agreename=1.

ii. Maiden name match: If Match_First = X and also either Match_Father_Surname= X or Match_Last_Surname=X, then set Name_BirthNameAgree=1.

iii. Name and Soundex: If any of these are true,

1. Match_First = X and Match_Last= N
2. Match_First = N and Match_Last= X
3. Match_First = X and either Match_Father_Surname = N or Match_Last_Surname = N
4. Match_First = N and either Match_Father_Surname = X or Match_Last_Surname =X

Then set Name_Close=1.

d. Demographics management: If Match_Race, Match_Marital, and Match_State_Birth all = X, then set Demographics=1.

e. Rare name pre-processing: Use the three files for female first names, male first names, , and last names.

i. Sort each of the files with male first names and female first names separately, by variable First, keeping variables First from each file and F_Percent and M_Percent respectively. Rename First to First_Name.

ii. Merge both sorted files on First, with resulting file contain variables: First, F_Percent, and M_Percent. (The percents are kept as distinct variables as some names can be either male or female first names: e.g., Leslie, Dana, Sandy, etc. and the distinct percents need to be maintained; i.e., some names will have both F_Percent and M_Percent filled, while most names will have only one of these two variables filled.)

iii. Sort the file with last names on Last, renaming Last to Last_Name.

iv. Merge the resulting file of first names from step 18.e.ii with the current temporary file on First_Name, to attach variables F_Percent and M_Percent.

v. Merge the resulting file of last names from step 18.e.iii with the current temporary file on Last_Name, to attach variables L_Percent.

f. Rare name management: Perform the next steps in order, creating the numeric variable RareName, which will take on values of missing or 1-6.

i. Missing names: If Last_Name or First_Name are missing/blank, then set RareName=.

ii. Source names not on reference files: Set RareName=6 for any of the following conditions:

1. L_Percent value cannot be determined from step 18.e.v.
2. Sex has the value of 1 or M and M_Percent cannot be determined from step 18.e.iv.
3. Sex has the value of 2 or F and F_percent cannot be determined from step 18.e.iv.

iii. Very common first and last names: Set RareName=1 for any of the following conditions:

1. L_Percent >= .0490491 and Sex has the value of 1 or M and M_Percent >= .0663340

2. L_Percent >= .0490491 and Sex has the value of 2 or F and F_Percent >=.0919882

iv. Common first and last names: Set RareName=2 for any of the following conditions:
1. L_Percent >= .0490491 and Sex has the value of 1 or M and M_Percent >= .0169621
2. L_Percent >= .0490491 and Sex has the value of 2 or F and F_Percent >=.0199908
3. L_Percent >= 0216809 and Sex has the value of 1 or M and M_Percent >= .0663340
4. L_Percent >= 0216809 and Sex has the value of 2 or F and F_Percent >=.0919882

v. Very rare first and last names: Set RareName=5 for any of the following conditions:
1. L_Percent <= .0024046 and Sex has the value of 1 or M and M_Percent <= .0018228
2. L_Percent <= .0024046 and Sex has the value of 2 or F and F_Percent <=.0019786

vi. Very rare first or last names: Set RareName=4 for any of the following conditions:
1. L_Percent <= .0024046 or both Sex has the value of 1 or M and M_Percent <= .0018228
2. L_Percent <= .0024046 or both Sex has the value of 2 or F and F_Percent <=.0019786

vii. Average name frequency: Set RareName=3 for any records not yet assigned a RareName value.

g. Assign Project Scores: Following this table, use Criteria in order to assign Project_Score as indicated.

Criteria	Project_Score
Agreement: Name, SSN, Gender and Date Of Birth Name_FullAgree=1 and SSN_Score=9 and Match_Sex=X and DOB_Agree=1	1
NDI Class variable=1 Class_Code=1	1.1
Agreement: Name, SSN, Gender and Month and Year Of Birth Name_FullAgree/agreename =1 and SSN_Score=9 and Match_Sex=X and DOB_MoYr/agreemoyr=1	2
Agreement: First Name, Maiden Name, SSN, Gender and Birth Name_BirthNameAgree/agreebirthname=1 and SSN_Score=9 and Match_Sex=X and DOB_Agree/agreebirth =1	3

Criteria	Project_Score
Agreement: First name, Maiden Name, SSN, Gender and Birth Month and Year Name_BirthNameAgree=1 and SSN_Score=9 and Match_Sex=X and DOB_MoYr/agreemoyr =1	3.1
Agreement: Parts of name, SSN, Gender and Birthdate Name_Close=1 and SSN_Score=9 and Match_Sex=X and DOB_Agree/agreebirth =1	3.2
Agreement: Parts of name, SSN, Gender and Month And Year Of Birth Name_Close=1 and SSN_Score=9 and Match_Sex=X and DOB_MoYr/agreemoyr =1	3.3
Agreement: Name, Gender, and Date of Birth; SSN with a typo Name_FullAgree=1 and Match_Sex=X and DOB_Agree/agreebirth=1 and SSN_Close=1	3.4
Agreement: Name, Gender, and Month and Year of Birth; SSN with a typo Name_FullAgree=1 and Match_Sex=X and DOB_MoYr =1 and SSN_Close=1	3.5
Agreement: First name, Maiden Name, Gender, Birthdate; SSN with a typo Name_BirthNameAgree=1 and Match_Sex=X and DOB_Agree=1 and SSN_Close=1	3.6
Agreement: First Name, Maiden Name, Gender, Month and Year of Birth; SSN with a typo Name_BirthNameAgree=1 and Match_Sex=X and DOB_MoYr =1 and SSN_Close=1	3.7
Agreement: Last name, First name, SSN with a typo and Gender and 2 out of 3 date of birth parts Name_FullAgree =1 and SSN_Close=1 and Match_Sex=X and DOB_2of3=1	4.1
Similar name, SSN with a typo/trans and Gender and 2 out of 3 date of birth parts Name_Close=1 and SSN_Close=1 and Match_Sex=X and DOB_2of3=1	4.2

Criteria	Project_Score
Agreement: Name, Gender and SSN Name_FullAgree=1 and Match_Sex=X and SSN_Score=9	4.5
Similar Name, Gender, and SSN Name_Close=1 and Match_Sex=X and SSN_Score=9	4.6
Agreement: Last name, SSN, Gender, 2 out of 3 date of birth parts Match_Last=X and SSN_Score=9 and Match_Sex=X and DOB_2of3=1	4.7
Agreement: Name, SSN, and date of birth Name_FullAgree=1 and SSN_Score=9 and DOB_Agree=1	4.9
Agreement: Name, SSN, and 2 out of 3 date of birth parts Name_FullAgree=1 and SSN_Score=9 and DOB_2of3=1	4.91
Agreement: Name, SSN with typo, and Date of Birth Name_FullAgree=1 and SSN_Close=1 and DOB_Agree=1	4.92
Agreement: SSN, DOB, and gender and agreement on either first name or last name SSN_Score=9 and DOB_Agree=1 and Match_Sex=X and (Match_First=X or Match_Last=X)	4.93
Agreement: SSN, DOB, and Gender and Demographics SSN_Score=9 and DOB_Agree=1 and Match_Sex=X and Demographics=1	4.94
SSN missing, Rare Name with agreement on: First Name, Last Name, Middle Initial, DOB, Demographics, and Gender SSN=missing/blank and RareName not either 1 or 2 and Name_FullAgree=1 and Match_Middle is either X or B and DOB_Agree=1 and Demographics=1 and Match_Sex=X	6.1

Criteria	Project_Score
SSN Missing, Rare Name, with agreement: First Name, Last Name, DOB, Demographics, and Gender SSN is missing/blank and RareName not either 1 or 2 and Name_FullAgree=1 and DOB_Agree=1 and Demographics=1 and Match_Sex=X	6.2
Agree on Name, Rare Name, DOB, Gender and Six+ Digits of SSN RareName not either 1 or 2 and (Name_FullAgree=1 or NameClose=1) and DOB_Agree=1 and Match_Sex=X and SSN_Score >= 6	7
Agree on First Name, Rare Name, Maiden Name, Gender and Date Of Birth, and Missing SSN Match_First=X and RareName not either 1 or 2 and (Match_Father_Surname=X or Match_Last_Surname=X) and Match_Sex=X and DOB_Agree=1 and SSN=missing/blank	8
SSN Missing, Very Rare Name, with Agreement: First Name, Last Name, Birth Year $\pm$ 3 Years, and Gender SSN=missing/blank and RareName one of 4, 5, or 6 and Name_FullAgree=1 and DOB_3Year=1 and Match_Sex=X	9
SSN Missing and Very Rare Name, With Agreement: First Name, Maiden Name/Fathers Surname, Birth Year $\pm$ 3 Years, and Gender SSN=missing/blank and RareName one of 4, 5, or 6 and Match_First=X and (Match_Father_Surname=X or Match_Last_Surname=X) and DOB_3Year=1 and Match_Sex=X	9.1
Agreement: SSN, Gender, (part of first name or part of last name or last name=father surname), and 2 out of 3 date of birth parts SSN_Score=9 and Match_Sex=X and (Match_First one of X or N or Match_Last one of X or N or Match_Last_Surname=X) and DOB_2of3=1	12
Agreement: SSN and first name SSN_Score=9 and Match_First one of X or N	12.1

Criteria	Project_Score
NDI-status=1 and NDI-Class=2 Status_Code=1 and Class_Code=2	13.2
NDI-status=1 and NDI-Class=3 Status_Code=1 and Class_Code=3	13.3
NDI-status=1 and NDI-Class=4 Status_Code=1 and Class_Code=4	13.4
SSN Missing and Agreement: First Name, Last Name, DOB, and Gender, Middle Initial not Mismatch with NDI Record SSN=missing/blank and Name_FullAgree=1 and DOB_Agree=1 and Match_Sex=X and Match_Middle not equal to missing/blank	14
Agreement: SSN and Gender SSN_Score=9 and Match_Sex=X	15
Rare Name and Agreement: First Name, Last Name, Middle Initial, DOB, Demographics, and Gender RareName not either 1 or 2 and Name_FullAgree=1 and Match_Middle one of B or X and DOB_Agree=1 and Demographics=1 and Match_Sex=X	16.1
Rare Name and Agreement: First Name, Last Name=Father Surname, DOB, and Gender RareName not either 1 or 2 and Match_First=X and Match_Last_Surname=X and DOB_Agree=1 and Match_Sex=X	16.2
Very Rare Name and Agreement: First Name, Last Name, Birth Year $\pm$ 3 Years, and Gender RareName one of 4, 5, or 6 and Name_FullAgree=1 and DOB_3Year =1 and Match_Sex=X	16.3

Criteria	Project_Score
Very Rare Name and Agreement: First Name, Last Name=Father Surname, Birth Year ± 3 Years, and Gender RareName one of 4, 5, or 6 and Match_First=X and Match_Last_Surname=X and DOB_3Year =1 and Match_Sex=X	16.4
Agreement: First Name, Last Name, DOB, and Gender Name_FullAgree=1 and DOB_Agree=1 and Match_Middle not equal to missing/blank and Match_Sex=X	16.5

19. Identify best match: Sort file by PatID, then by Project_Score, then by *descending* Prob_Score. Per PatID, assign values as follows for distinct records by Project_Score and Prob_Score:
    a. Algorithm3=1 to first record identified.
        i. If more than one record has identical minimum value for Project_Score and maximum value for Prob_Score, write both records to **Potential_Ties**.
    b. Algorithm3=0 to all other records that contain a Project_Score from 1-16.5.
    c. Algorithm3=.S (special missing value) to all other records that contain a missing Project_Score.
    d. Write all of these records to **NDI_Selection**.
20. As some records may not have been processed through each of all 3 Algorithms and written to **NDI_Selection**, set Algorithm1, Algorithm2 and Algorithm3 to .S (special missing value), for any values that are not either 1 or 0.
21. Complete the **NDI_Selection, Potential_Ties** and **Duplicates** files, following the data dictionary in this section. Sort each by PatID and save all 3 files to the DPLocal data library established for this package.
22. As a quality control check, for each of the 3 algorithms, sum the count of records in **NDI_Selection** with Algorithm#= 1 or 0, the count of records in **Potential_Ties** with Algorithm#= 1 or 0, and the count of all records in the **Duplicates** file. The sum of the three counts should be equal to the number of records in the **Combined** file.

V. Aggregate of NDI Quality Control File

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
NDI_Source	Char(3)	$3.	NDI source file code	M= Match file N= NoMatch file R= Reject file MN= Found in the Match and NoMatch files MR= Found in the Match and Reject files NR= Found in the NoMatch and Reject files MNR= Found in the Match, NoMatch, and Reject files NF= Not found in any NDI file	Code for source of NDI file
Link	Char(1)	$1	Link code	B= Found in both NDI and health plan files N= Found only in NDI files H= Found only in health plan file	Quality control indicator of links between NDI returned files and health plan source files
Count	Num(8)	comma10.	Count	0+	Count of records

Methods for creation of aggregation of NDI quality control file

1. Use the patient-level file **NDI_Quality_Control** as saved in the DPLocal data library established for this package.
2. Set DPID_SiteID to the value for the health plan.

3. Examine the **NDI_Quality_Control** file and if on the basis of the combination of PatID/Sequence_Number, a record is found in more than one NDI file, then this should be resolved to a single row, setting the variable NDI_Source, following these rules:
    a. MN= Found in the Match and NoMatch files
    b. MR= Found in the Match and Reject files
    c. NR= Found in the NoMatch and Reject files
    d. MNR= Found in the Match, NoMatch, and Reject file
4. Aggregate the file on variables NDI_Source and Link. Set Count with the resulting counts.
5. Sort the file by NDI_Source and Link, name it as ***DPID_SiteID*_NDI_Quality_Control** and save to the MSOC data library.

VI. Aggregate of NDI Returns Pre and Post

This file will be returned to the[Lead Analytic Site]. It will contain aggregate counts of returned NDI records selected by identification of possible health plan filtering of records, and will follow the data dictionary shown below.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
FileInd	Char(5)	$5.	File indicator	PRE= NDI_Outbound_Records (Package 4) POST = NDI_Quality_Control	Indicator of file used for counts; "post" may have had filtering for patients no longer eligible for research
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Count	Num(8)	comma10.	Count	0+	Count of records

Methods for creation of aggregation of selection of NDI records

1. Use the file **NDI_Outbound_Records** as found in the DPLocal data library from Package 4. Determining availability of this file will have already occurred in Section I. Keep rows with distinct values of variable PatID.
2. Use the file **All_Cohorts**, found in the DPLocal data library for Package 1. Availability of this file will have already occurred in Section I. Keep variables PatID and CohortInd.
3. On PatID, link the file from step 2 with **All_Cohorts**, together, keeping rows from found in the step 2 **NDI_Outbound_Records** only.
    a. Assign FileInd="PRE" for these records.
4. Use the file **NDI_Quality_Control** created in Section I Keep only rows where Link=B or H and resolve to distinct rows by PatID.
    a. On PatID, link the resulting file from step 4 with **All_Cohorts**, to obtain variable CohortInd.
    b. Assign FileInd="POST" for these records.

5. Determine if the health plan identifies a file that lists PatIDs not eligible for research.
    a. If so, then on PatID, merge the subset **NDI_Quality_Control** file from step 4 with such file, keeping all rows from the step 4 subset **NDI_Quality_Control** file, containing PatIDs not found in the health plan file that lists PatIDs not eligible for research. Save this file temporarily as **Post**; continue with step 8.
    b. Otherwise, copy the subset **NDI_Quality_Control** file to temporary file **Post** and continue with next steps.
6. Concatenate (SQL union) the file from step 3 (derivative of **NDI_Outbound_Records**) with the file from step 5 (**Post**).
7. Aggregate the file from step 6 on variables FileInd and CohortInd. Set Count with the resulting counts.
8. Set DPID_SiteID to the value for the health plan.
9. Sort the file by FileInd and CohortInd, name it as **NDI_Returned_Results** and save to the MSOC data library.

VII. Aggregate of NDI Descriptives

This file will be returned to the[Lead Analytic Site]It will contain aggregate data of NDI matching variables and will follow the data dictionary shown below. It will be used to create display table found in Appendix D.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(3)	$3.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort All= Both cohorts	
Category	Char(25)	$25.	Column categories	Status code= 1 Status code= 0 Status code=1 or 0 Death only Death and Disenrolled Disenrolled only LTF All	Subsets by NDI status codes or by reasons submitted to NDI
Factor	Char(15)	$15.	Row categories	See Appendix C	Factors that define rows
Pats_Match_Return	Num(8)	Comma10.	Patients with at least 1 match returned	0+	Count of unique patients (PatIDs) from the Combined file

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Pats_Match_Avg	Num(8)	6.2	Per Patient: Avg # of matches	1+	Mean number of NDI records per patient (PatID) submitted
Pats_Match_Min	Num(3)	1.	Per Patient: Min # of matches	1+	Minimum number of NDI records per patient (PatID) submitted
Pats_Match_Max	Num(3)	Comma10.	Per Patient: Max # of matches	1+	Maximum number of NDI records per patient (PatID) submitted
Pats_Death_Complete	Num(8)	Comma10.	Patients with full death date	0+	Count of unique patients (PatIDs) with complete death date returned from NDI
Pats_Death_Month	Num(8)	Comma10.	Patients with filled death date month	0+	Count of unique patients (PatIDs) with filled death date month returned from NDI
Pats_Death_Day	Num(8)	Comma10.	Patients with filled death date day	0+	Count of unique patients (PatIDs) with filled death date day returned from NDI
Pats_Death_Year	Num(8)	Comma10.	Patients with filled death date year	0+	Count of unique patients (PatIDs) with filled death date year returned from NDI
CountMatch	Num(8)	Comma10.	Count of matches	0+	Count of patients matched, for Category by Factor
Percent	Num(8)	Percent7.2	Percent	0+	Percent = CountMatch / Pat_Match_Return

Methods for creation of aggregation of selection NDI matching factors

1. Use the patient-level file **NDI_Selection** as saved in the DPLocal data library established for this project for all values except for those indicated as requiring use of file **Combined_Distinct** or **Potential_Ties**.
2. Run through all steps from 3 through 12 for each value of CohortInd in the data dictionary as follows:
    a. Subset rows for and set CohortInd="C1".
    b. Subset rows for and set CohortInd="C2".

   c. Include all rows and set CohortInd="ALL".

3. For each of steps 4 through 12, set Category to the values found in the data dictionary and then create aggregates for each level of Category, using the following definitions:
   a. Status code= 1, when variable Status_Code=1.
   b. Status code= 0, when variable Status_Code=0.
   c. Status code=1 or 0, when variable Status_Code=1 or =0.
   d. Death only, when Reason_Select=DE.
   e. Death and Disenrolled, when Reason_Select=B.
   f. Disenrolled only, when Reason_Select=DI.
   g. LTF, when Reason_Select=LTF.
   h. All, when no filter/subset is applied.

For steps 4 through 11, use file **Combined_Distinct**, from Section IV.

4. Set Pats_Match_Return to the number of unique PatIDs.
5. Set Pats_Match_Avg to the mean number of rows per PatID
6. Set Pats_Match_Min to the minimum number of rows per PatID
7. Set Pats_Match_Max to the maximum number of rows per PatID
8. Set Pats_Death_Complete to the number of records from NDI with a complete NDI_Death_Date NOTE: For this and the next 3 variables we are counting per record and not by PatID, as the NDI may return more than one record per PatID with different values.
9. Set Pats_Death_Month to the number of records from NDI with a filled NDI_Death_MO.
10. Set Pats_Death_Day to the number of records from NDI with a filled NDI_Death_DAY.
11. Set Pats_Death_Year to the number of records from NDI with a filled NDI_Death_YR.
12. Create aggregates for each of the Factors as defined in Appendix C, following the definition indicated:
    a. Set the Factor variable, concatenating Appendix C Sort_Order_VII, a hyphen, and the Appendix C Factor.
    b. Set CountMatch for each Category and Factor, as the number of distinct PatIDs identified.
    c. Set Percent for each Category and Factor as the CountMatch (step 12.b) divided by Pat_Match_Return (step 4) for the Category, multiplied by 100, rounded to 2 decimals. If Pat_Match_Return (denominator) is missing, then set Percent=0.
13. Set DPID_SiteID to the value for the health plan.
14. Assemble all data into a single table following the data dictionary above. Sort by Factor, CohortInd, and Category, name the file **NDI_Descriptives** and save to the MSOC data library for this package.

VIII. Aggregate Comparison of Death Sources

This file will be returned to the [Lead Analytic Site]. It enables comparison of death sources between health plans and NDI ascertainment of death. This is the dataset that will be returned to [Lead Analytic Site]. It will be used to create display table found in Appendix E.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(3)	$3.	Cohort ID	C1=Exposed cohort C2=Baseline cohort ALL=Both cohorts	
Algorithm_Number	Num(3)	1.	Algorithm number	1, 2, or 3	Identifies the algorithm used to create matches from NDI records
Source	Char(1)	1.	Source of death information	L = Other, locally defined N = National Death Index S = State Death files T = Tumor data A = All	Each of these distinct; not crossed with Confidence
Confidence	Char(1)	1.	Confidence that the patient drawn from the Source data represents the actual patient (contrasts with Confidence in the Cause of Death table).	E = Excellent F = Fair P = Poor A = All	Each of these distinct; not crossed with Source

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Factor	Char(20)	$20.	Row categories	See Appendix C	Factors that define rows
True_Pos	Num(8)	comma10.	Count (A)	0+	Count of death in health plan and death in NDI
False_Pos	Num(8)	comma10.	Count (B)	0+	Count of alive in health plan and death in NDI
False_Neg	Num(8)	comma10.	Count (C)	0+	Count of death in health plan and alive in NDI
True_Neg	Num(8)	comma10.	Count (D)	0+	Count of alive in health plan and alive in NDI
Test_Positive	Num(8)	comma10.	Count (A +B)	0+	Count of True_Pos (A) + False_Pos (B)
Test_Negative	Num(8)	comma10.	Count (C +D)	0+	Count of False_Neg (C) + True_Neg (D)
Plan_Death	Num(8)	comma10.	Count (A + C)	0+	Count of True_Pos (A) + False_Neg (C)
Plan_NoDeath	Num(8)	comma10.	Count (B +D)	0+	Count of False_Pos (B) + True_Neg (D)
Total_Pop	Num(8)	comma10.	Count of population	0+	Count of True_Pos (A) + False_Pos (B) + False_Neg (C) + True_Neg (D)
Sensitivity	Num(8)	5.2	Sensitivity	000.00% +	True_Pos (A) divided by [(True_Pos (A) + False_Neg (C)]
Specificity	Num(8)	5.2	Specificity	000.00% +	True_Neg (D) divided by [False_Neg (C) + True_Neg (D)]
PPV	Num(8)	5.2	Positive Predictive Value	000.00% +	True_Pos (A) divided by [True_Pos (A) + False_Pos (B)]

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
NPV	Num(8)	5.2	Negative Predictive Value	000.00% +	True_Neg (D) divided by [False_Neg (C) + True_Neg (D)]

Methods for creation of aggregate of matching to Data Partner death data

1. Using the table referenced by Appendix B, identify the row with a matching DPID_SiteID to that of the health plan executing the package.
    a. If DeathTable="Y", check that the file named **Death_NDI** is found in the DPLocal data library for Package 1.
    b. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 1.
    c. Otherwise if found, proceed with all other processes.
    d. Save the LastDeathDate value from this row to a temporary variable LastDeathDate that can be made available to a later step.
2. If the Data Partner does not have a Death table (i.e., DeathTable="N"), create a pseudo-death table, using the Encounter_NDI dataset, saved in the DPLocal data library from package #1:
    a. Filter for all rows where EncType=IP and Discharge_Status="EX". This will create a list of encounters where the patient died during an inpatient encounter. Keep only variable PatID. Set Confidence="A" and Source="A" for all rows. This will become a surrogate Death table and will be used below.
    b. Create a temporary variable, called NDI_EndDate and for this project, set this date to 12/31/18.
    c. Set LastDeathDate as follows:
        i. If DPMaxDate (see Section IV, Step 2.b.ii.) is earlier than NDI_EndDate, then set LastDeathDate to DPMaxDate.
        ii. If NDI end date is earlier than DPMaxDate, then set LastDeathDate to NDI_EndDate.
3. Use the patient-level file **NDI_Selection** created earlier in Section IV. Run through all steps from 5 through 13 for each value of CohortInd, Source, and Confidence in the data dictionary. This will result in 3 X 9 progressions, minus 1 (26 progressions) through the steps, as determined by:
    a. 3 levels of CohortInd by:
    b. 5 individual levels of Source, for Confidence="A".
    c. 3 individual levels of Confidence for Source="A".
    d. 1 level of Source="A" and Confidence="A". Note that we do not need crossings of all of the individual levels of Source by all of the individual levels of Confidence, other than specified above. This table shows all possible crossings required.

#	CohortInd	Source	Confidence
1	C1	L	A
2	C1	N	A
3	C1	S	A
4	C1	T	A
5	C1	A	A
6	C2	L	A
7	C2	N	A
8	C2	S	A
9	C2	T	A
10	C2	A	A
11	ALL	L	A
12	ALL	N	A
13	ALL	S	A
14	ALL	T	A
15	C1	A	E
16	C1	A	F
17	C1	A	P
18	C2	A	E
19	C2	A	F
20	C2	A	P
21	ALL	A	E

#	CohortInd	Source	Confidence
22	ALL	A	F
23	ALL	A	P
24	ALL	A	A

4. Scan the Death table from step 1.a or the pseudo Death table from step 2.a for all possible combinations from the table above for Source and Confidence. Retain combinations only when there is at least one row with a specific value for Source and at least one row with a specific value of Confidence. (There will always be the 3 values of CohortInd of C1, C2, and ALL.) Proceed with all steps below for only the resulting combinations in the table above. For example, if a Death table has values of Source=L & S only, along with Confidence=E only, then the only combinations from the table above that can be calculated are indicated by row numbers: 1, 3, 5-6, 8, 10-11, 13, 15, 18, 21, & 24.
5. Proceed with all steps from 6 through 14 for each Algorithm 1 through Algorithm 3 in **NDI_Selection**.
    a. Set Algorithm= the algorithm number and subset of records.
6. Use the **NDI_Outbound_Records** file from Package 4 DPLocal data library. (Availability of this file has already been determined in Section I) Keep only variable PatID with distinct values. On PatID, link with the filtered plan's death table from step 3 above. Set a new variable HP_Death as follows:
    a. If the PatID is found in the filtered Death table, then set to "Dead".
    b. If the PatID is not found in the Death table, then set to "Alive".
7. Summarize the rows of **NDI_Selection,** by PatID, setting Algorithm# to the highest value of 1 or 0. Doing so will identify if any health record matched to any NDI record (value of 1) or no HP record matched to a NDI record (value of 0). The result will be a distinct row per PatID. Ensure that comparisons of health plan death data with NDI death data are truncated at the last date for which the health plan is believed to have death data completeness. Filter the file created at this step for only those rows where NDI_Death_Date is equal to LastDeathDate or earlier
8. Proceed with all steps from 9 through 14, for each Factor in Appendix C with Sort_Order_VIII not equal to "N/A".
9. Filter the file created in step 7 to include only rows where the PatID meets the Factor definition. Create the factor variable, concatenating Appendix C Sort_Order_VIII, a hyphen, and the Appendix C Factor.
10. Link the file from step 6 with the filtered **NDI_Selection** file from step 7, keeping all rows from **NDI_Outbound_Records** whether the PatID is found in the filtered **NDI_Selection** table or not.
    a. Create a new variable NDI_Death and set as follows:
        i. If Algorithm#=1, then set to "Dead".
        ii. If Algorithm#=0, then set to "Alive".

   b. Note: if a PatID is not found in the filtered **NDI_Selection** file, set NDI_Death to "Alive"

11. Create counts by combinations of DP_DeathData and NDI_Death as follows:
    a. True_Pos = count of rows where HP_Death="Dead" and NDI_Death="Dead".
    b. False_Pos = count of rows where HP_Death="Alive" and NDI_Death="Dead".
    c. False_Neg = count of rows where HP_Death="Dead" and NDI_Death="Alive".
    d. True_Neg = count of rows where HP_Death="Alive" and NDI_Death="Alive".
12. Summarize counts of created variables in step 11 and set as follows:
    a. Test_Positive = True_Pos + False_Pos.
    b. Test_Negative = False_Neg + True_Neg
    c. Plan_Death = True_Pos + False_Neg
    d. Plan_NoDeath = False_Pos + True_Neg
    e. Total_Pop = True_Pos + False_Pos + False_Neg + True_Neg
        i. As a QC check, Total_Pop should be equal for every row within CohortInd
13. Calculate percents:
    a. Sensitivity = True_Pos divided by (True_Pos + False_Neg), multiply by 100 and round to 2 decimal places.
    b. Specificity = True_Neg divided by (False_Pos + True_Neg), multiply by 100 and round to 2 decimal places.
    c. PPV = True_Pos divided by (True_Pos + False_Pos), multiply by 100 and round to 2 decimal places.
    d. NPV = True_Neg divided by (False_Neg + True_Neg), multiply by 100 and round to 2 decimal places.
14. Fill the data dictionary above with the values for all variables.
15. Sort the table by CohortInd, Algorithm, and Factor. Name the table **NDI_Death_Compare** and save to the MSOC data library

Appendix A – State Postal Codes Accepted by NDI

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Alabama	AL	01	New Jersey	NJ	31
Alaska	AK	02	New Mexico	NM	32
Arizona	AZ	03	New York	NY	33
Arkansas	AR	04	North Carolina	NC	34
California	CA	05	North Dakota	ND	35
Colorado	CO	06	Ohio	OH	36
Connecticut	CT	07	Oklahoma	OK	37
Delaware	DE	08	Oregon	OR	38
District of Columbia	DC	09	Pennsylvania	PA	39
Florida	FL	10	Rhode Island	RI	40
Georgia	GA	11	South Carolina	SC	41
Hawaii	HI	12	South Dakota	SD	42
Idaho	ID	13	Tennessee	TN	43
Illinois	IL	14	Texas	TX	44
Indiana	IN	15	Utah	UT	45
Iowa	IA	16	Vermont	VT	46
Kansas	KS	17	Virginia	VA	47

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Kentucky	KY	18	Washington	WA	48
Louisiana	LA	19	West Virginia	WV	49
Maine	ME	20	Wisconsin	WI	50
Maryland	MD	21	Wyoming	WY	51
Massachusetts	MA	22	Puerto Rico	PR	52
Michigan	MI	23	Virgin Islands	VI	53
Minnesota	MN	24	Guam	GU	54
Mississippi	MS	25	Canada	CN	55
Missouri	MO	26	Cuba	CU	56
Montana	MT	27	Mexico	MX	57
Nebraska	NE	28	Remainder of World	RW	59
Nevada	NV	29	Unknown		99
New Hampshire	NH	30			

Appendix B – Death Data Completeness Input File

[Lead Analytic Site] will create this input file based on review of health plans' QA data from the ETL used to run Package 1.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
LastDeathDate	Num(4)	mmddyy10.	Final death date within truncation	Valid SAS dates	
DeathTable	Char(1)	$1.	Death table available	Y=Yes, has a Death table N=No, has no Death table	

Sample Rows

DPID_SiteID	LastDeathDate	DeathTable
Health Plan 1		N
Health Plan 2	x/xx/xxxx	Y
Health Plan 3	x/xx/xxxx	Y
Health Plan 4	x/xx/xxxx	Y

Appendix C – Factors Defining Aggregates in Sections VII and VIII

Sort_Order_VII	Sort_Order_VIII	Factor	Definition[5]
01	N/A	Status_Code_1	Status_Code=1
02	N/A	Status_Code_0	Status_Code=0
03	04[6]	Exact_Num	# of distinct PatIDs containing *any* row with Exact="*"
04	05	Class_1_total	# of distinct PatIDs with Class_Code=1
05	N/A	Class_1_ties	# of distinct PatIDs with Class_Code=1, found in **Potential_Ties** table
06	06	Class_2_total	# of distinct PatIDs with Class_Code=2
07	N/A	Class_2_ties	# of distinct PatIDs with Class_Code=2, found in **Potential_Ties** table
08	07	Class_2_LT_34.5	# of distinct PatIDs with Class_Code=2 and Prob_Score less than 34.5
09	08	Class_2_GE_34.5	# of distinct PatIDs with Class_Code=2 and Prob_Score greater than or equal to 34.5
10	09	Class_2_GE_39.5	# of distinct PatIDs with Class_Code=2 and Prob_Score greater than or equal to 39.5
11	10	Class_2_GE_44.5	# of distinct PatIDs with Class_Code=2 and Prob_Score greater than or equal to 44.5
12	11	Class_2_GE_49.5	# of distinct PatIDs with Class_Code=2 and Prob_Score greater than or equal to 49.5

[5] Factors in this table are not necessarily mutually exclusive. For example, a record satisfying the Definition for Factor "Class_3_GE_37.5" will also satisfy the Definition for Factor "Class_3_GE_32.5"; a record that satisfies the Definition for Factor "Algorithm_2_COD" also satisfies the Definition for Factor "Algorithm_2".

[6] For Section VIII, Sort_Orders 04-27 repeat for each Algorithm. Thus, for Algorithm 1, the Sort_Order becomes: 01, 04-27, for Algorithm 2, the Sort_Order becomes 02, 04-27, and for Algorithm 3, the Sort_Order becomes 03, 04-27.

Sort_Order_VII	Sort_Order_VIII	Factor	Definition[5]
13	12	Class_2_GE.54.5	# of distinct PatIDs with Class_Code=2 and Prob_Score greater than or equal to 54.5
14	13	Class_3_total	# of distinct PatIDs with Class_Code=3
15	N/A	Class_3_ties	# of distinct PatIDs with Class_Code=3, found in **Potential_Ties** table
16	14	Class_3_LT_27.5	# of distinct PatIDs with Class_Code=3 and Prob_Score less than 27.5
17	15	Class_3_GE_27.5	# of distinct PatIDs with Class_Code=3 and Prob_Score greater than or equal to 27.5
18	16	Class_3_GE_32.5	# of distinct PatIDs with Class_Code=3 and Prob_Score greater than or equal to 32.5
19	17	Class_3_GE_37.5	# of distinct PatIDs with Class_Code=3 and Prob_Score greater than or equal to 37.5
20	18	Class_3_GE_42.5	# of distinct PatIDs with Class_Code=3 and Prob_Score greater than or equal to 42.5
21	19	Class_3_GE_47.5	# of distinct PatIDs with Class_Code=3 and Prob_Score greater than or equal to 47.5
22	20	Class_4_total	# of distinct PatIDs with Class_Code=4
23	N/A	Class_4_ties	# of distinct PatIDs with Class_Code=4, found in **Potential_Ties** table
24	21	Class_4_LT_22.5	# of distinct PatIDs with Class_Code=4 and Prob_Score less than 22.5
25	22	Class_4_GE_22.5	# of distinct PatIDs with Class_Code=4 and Prob_Score greater than or equal to 22.5
26	23	Class_4_GE_27	# of distinct PatIDs with Class_Code=4 and Prob_Score greater than or equal to 27.0
27	24	Class_4_GE_32.5	# of distinct PatIDs with Class_Code=4 and Prob_Score greater than or equal to 32.5
28	25	Class_4_GE_37	# of distinct PatIDs with Class_Code=4 and Prob_Score greater than or equal to 37.0
29	26	Class_4_GE_42.5	# of distinct PatIDs with Class_Code=4 and Prob_Score greater than or equal to 42.5
30	27	Class_5_total	# of distinct PatIDs with Class_Code=5
31	N/A	Class_5_ties	# of distinct PatIDs with Class_Code=1, found in **Potential_Ties** table
32	01	Algorithm_1_Saved	# of distinct PatIDs saved for Algorithm 1 with no ties, where Algorithm1=1

Sort_Order_VII	Sort_Order_VIII	Factor	Definition[5]
33	N/A	Algorithm_1_Ties	# of distinct PatIDs saved for Algorithm 1 with ties (From file **Potential_Ties**)
34	N/A	Algorithm_1_COD	# of distinct PatIDs saved for Algorithm 1 with COD_Available=1, where Algorithm1=1
35	02	Algorithm_2_Saved	# of distinct PatIDs saved for Algorithm 2 with no ties, where Algorithm2=1
36	N/A	Algorithm_2_Ties	# of distinct PatIDs saved for Algorithm 2 with ties (From file **Potential_Ties**)
37	N/A	Algorithm_2_COD	# of distinct PatIDs saved for Algorithm 2 with COD_Available=1, where Algorithm2=1
38	03	Algorithm_3_Saved	# of distinct PatIDs saved for Algorithm 3 with no ties, where Algorithm3=1
39	N/A	Algorithm_3_Ties	# of distinct PatIDs saved for Algorithm 3 with ties (From file **Potential_Ties**)
40	N/A	Algorithm_3_COD	# of distinct PatIDs saved for Algorithm 3 with COD_Available=1

Appendix D – Display Table Sample for Aggregate of NDI Descriptives, Section VII

	Status code= 1		Status code= 0		Status code=1 or 0		Death only		Death and Disenrolled		Disenrolled only		LTF		All (total submitted to NDI)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Patients with at least 1 match returned	9,999	100.0 %	9,999	100.0 %	9,999	100.0%	9,999	100.0 %	9,999	100.0%	9,999	100.0 %	9,999	100.0 %	9,999	100.0 %
Per Patient: Avg # of matches	9.99	-	9.99	-	9.99	-	9.99	-	9.99	-	9.99	-	9.99	-	9.99	-
Per Patient: Min # of matches	9	-	9	-	9	-	9	-	9	-	9	-	9	-	9	-
Per Patient: Max # of matches	9	-	9	-	9	-	9	-	9	-	9	-	9	-	9	-
Status Code=1	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Status Code=0	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Exact	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 1 total	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 1 # of ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 2 total	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 2 # of ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
<34.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥34.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥39.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥44.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥49.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥54.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 3 total	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%

	Status code= 1		Status code= 0		Status code=1 or 0		Death only		Death and Disenrolled		Disenrolled only		LTF		All (total submitted to NDI)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Class 3 # of ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
<27.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥27.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥32.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥37.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥42.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥47.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 4 total	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 4 # of ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
<22.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥22.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥27.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥32.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥37.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
≥42.5	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 5 total	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Class 5 # of ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 1	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 1 with ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%

	Status code= 1		Status code= 0		Status code=1 or 0		Death only		Death and Disenrolled		Disenrolled only		LTF		All (total submitted to NDI)	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Algorithm 1 with COD available	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 2	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 2 with ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 2 with COD available	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 3	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 3 with ties	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%
Algorithm 3 with COD available	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%	999	99.9%

Appendix E – Display Table Sample for Aggregate Comparison of Death Sources, Section VII

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Algorithm 1 Total Without Ties	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Exact	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 1 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 2 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥39.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥44.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥49.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥54.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 3 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Algorithm 2 Total Without Ties	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Exact	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 1 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 2 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥39.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥44.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥49.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥54.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 3 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
<27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Algorithm 3 Total Without Ties	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Exact	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 1 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 2 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
≥34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥39.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥44.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥49.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥54.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 3 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

Package 6: Save the best NDI match, and create files to be used in final study analyses

Overview

This Program Package saves the 'best' NDI match, and create files to be used in final study analyses.

Using results returned by the NDI, a user patient record may match none, one, or multiple NDI death records. The purpose of this program is to:

- Create filtering logic that will be applied to the NDI patient level data held by a site (i.e., health plan), saving per patient, one and only one NDI death record where a match occurred
- Create death and cause of death patient-level files at the sites, suitable for analysis
- Create multiple aggregate data files, returned to the [list lead analytic site] for evaluation by the study team

I. Death Table from Saved Match Records

The following is the data dictionary of patient-level data that will be saved by the processes in this section. This file is a modified version of the SCDM v7.0.0 Death table structure. A PatID is represented on one and only one row.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char(#)	Site specific length	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
DeathDt	Num(4)	mmddyy10.	NDI Date of death	Valid SAS date	As appears on the NDI record
DtImpute	Char(1)	$1.	Imputed death date	N = Not imputed	Defaulted to Not Imputed
Source	Char(1)	$1.	Source of death information	N = National Death Index	Defaulted to National Death Index
Confidence	Char(1)	$1.	Confidence	E = Excellent	Defaulted to Excellent
State_Death_Code	Char(3)	$3.	State of death code	See Appendix A	
Death_Cert	Char(6)	$6.	Death certificate number	Free text	Jurisdiction death certificate number

Methods to create the patient-level Death Table from Saved Match Records

1. Check that the file named **NDI_Selection** is found in the DPLocal data library from Package 5.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal library.
    b. Otherwise if found, proceed with all other processes.
2. Filter the file **NDI_Selection** based on the following logical expression and save as temporary file **NDI_Saved_Matches_Temp1**, keeping variables PatID, State_Death_Code, Death_Cert, NDI_Death_Date and CohortInd.
    a. Logical expression: Algorithm1=1.
    b. Write the logical expression in step 2.a into the signature file for this package.

3. Determine if the health plan identifies a file that lists PatIDs excluded from research.
    a. If so, then on PatID, merge the **NDI_Saved_Matches_Temp1** file with such file, keeping all rows from **NDI_Saved_Matches_Temp1** containing PatIDs *not found* in the health plan file that lists PatIDs excluded from research. Save this file temporarily as **NDI_Saved_Matches_Temp2**; continue with step 4.
    b. Otherwise, copy **NDI_Saved_Matches_Temp1** to temporary file **NDI_Saved_Matches_Temp2** and continue with next steps.
4. Modify variables as follows:
    a. Rename NDI_Death_Date to DeathDt.
    b. For DtImpute, set value to N for all rows.
    c. For Source, set value to N for all rows.
    d. For Confidence, set value to E for all rows
5. Populate all variables shown in the data dictionary in this section.
6. Sort by PatID and save as **NDI_Death** to the DPLocal data library for this package.
7. Save both temporary files **NDI_Saved_Matches_Temp1** and **NDI_Saved_Matches_Temp2** for use in a subsequent section.

II. Cause of Death Table from Saved Match Records

The following is the data dictionary of patient-level data that will be saved by the processes in this section. This file is a modified version of the SCDM v7.0.0 Cause of Death table structure. A PatID can be represented on more than one row.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char(#)	Site specific length	Patient ID	Alpha-Numeric	#= Site specific length
COD	Char(8)	$8.	Cause of death	Free text	
CodeType	Char(2)	$2.	Code type	09 = ICD-9 10 = ICD-10	
CauseType	Char(3)	$3.	Cause type	C = Contributory I = Immediate/Primary O= Other R=Record U = Underlying 358 = Underlying coded into 358 groups 113 = Underlying coded into 113 groups 130 = Underlying coded into 130 groups for infants	Values R, 358, 113, & 130 are enhancements

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
CauseCat	Char(1)	$1.	Cause category	E=Entity R=Record U=Any underlying	This variable is an enhancement; Any underlying (U) be CauseType=U, 358, 113, or 130
Position1	Char(1)	$1.	Cause position	Part/line number on death certificate 1=Part I, line 1 (a) 2=Part I, line 2 (b) 3=Part I, line 3 (c) 4=Part I, line 4 (d) 5=Part I, line 5 (e) 6=Part II R=Record cause category U=Any underlying	This variable is an enhancement; 1-6, when CauseCat=E R, when CauseCat=R Any underlying (U) can be CauseType=U, 358, 113, or 130
Position2	Num(3)	1.	Position Sequence	1-7 or missing	Sequence of condition within part/line code
Cause_Sequence	Num(3)	2.	Sequence Number	1-20 or missing	Up to 20 cause of death codes are available for each of entity COD and record COD
Nature_Injury	Char(1)	$1.	Nature of Injury Code	1= ICD–9 nature of Injury code 0=All other codes N=Not applicable	Value of 1 is permitted only when CodeType=09 N is for Entity codes only
Source	Char(1)	$1.	Source	N = National Death Index	Defaulted to National Death Index

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Confidence	Char(1)	$2.	Confidence	E=Excellent	Defaulted to Excellent

Methods to create the patient-level Cause of Death Table from Saved Match Records

1. Check that the file named **NDI_Cause** is found in the DPLocal data library from Package 5.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal library.
    b. Otherwise if found, proceed with all other processes.
2. Using the **NDI_Death** file from Section I in this package, on PatID, State_Death_Code, and Death_Cert, link with the **NDI_Cause** file, that had been written to the DPLocal data library from Package 5, to obtain variables Cause_Underlying, Cause_Recode_358, Cause_Recode_113, Cause_Recode_130, Number_Axis, EntityPosition1 – EntityPosition20, EntitySequence1 – EntitySequence20, EntityCode1 – EntityCode20, EntityInjury1 – EntityInjury20, Number_Record, RecordCode1 – RecordCode20, and RecordInjury1 – RecordInjury20.
    a. Note that some rows from **NDI_Death** may not be found in **NDI_Cause**. Use only rows found in **NDI_Cause**.
    b. In case of duplicates in **NDI_Cause** where multiple rows have identical values for all variables (i.e., "pure" duplicates), retain only a single row.
3. Create a series of rows for the present table, populating each with PatID, CodeType=10, Source=N, and Confidence=E. Follow these rules for creating distinct rows per PatID and the cause variables from **NDI_Cause**:
    a. If Cause_Underlying is filled, create a row where COD=Cause_Underlying, CauseType=U, CauseCat=U, Position1=U, Position2=missing, Cause_Sequence=missing, and Nature_Injury=N.
    b. If Cause_Recode_358 is filled, then create a row where COD=Cause_Recode_358, CauseType=358, CauseCat=U, Position1=U, Position2=missing, Cause_Sequence=missing, and Nature_Injury=N.
    c. If Cause_Recode_113 is filled, then create a row where COD=Cause_Recode_113, CauseType=113, CauseCat=U, Position1=U, Position2=missing, Cause_Sequence=missing, and Nature_Injury=N.
    d. If Cause_Recode_130 is filled, then create a row where COD=Cause_Recode_130, CauseType=130, CauseCat=U, Position1=U, Position2=missing, Cause_Sequence=missing, and Nature_Injury=N.
    e. Create up to 20 rows, determined by the value of Number_Axis, following these directives, where "#" is a counter from 1 up to the value of Number_Axis: COD=EntityCode#, CauseCat=E, Position1=EntityPosition#, Position2= EntitySequence#, Cause_Sequence=#, and Nature_Injury=N. Set CauseType as follows:
        i. If Position1=1, set CauseType=I.

ii. If Position1=2 through 5, set CauseType=O.
iii. If Position1=6, set CauseType=C.
iv. If Position1 is not between 1 and 6, set CauseType=O.

f. Create up to 20 more rows, determined by the value of Number_Record, following these directives, where "#" is a counter from 1 up to the value of Number_Record: COD=RecordCode#, CauseType=R, CauseCat=R, Position1=R, Position2= missing, Cause_Sequence=#, and Nature_Injury= RecordInjury#.

4. Populate all variables shown in the data dictionary in this section.
5. Sort by PatID and save as **NDI_COD** to the DPLocal data library for this package.

III. Aggregate of NDI Saved Matches Pre and Post

This file will be returned to [Lead Analytic Site]. It will contain aggregate counts of saved NDI records, and will follow the data dictionary shown below.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
FileInd	Char(4)	$4.	File indicator	PRE=NDI_Selection All POST=NDI_Selection Eligible	Indicator of file used for counts; "post" may have had filtering for patients no longer eligible for research
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Count	Num(8)	comma10.	Count	0+	Count of records

Methods for creation of aggregation of selection of NDI saved match records

10. Use the temporary files **NDI_Saved_Matches_Temp1** and **NDI_Saved_Matches_Temp2** created in an earlier section. Assign FileInd as follows:
    a. PRE for all rows from **NDI_Saved_Matches_Temp1**
    b. POST for all rows from **NDI_Saved_Matches_Temp2**
11. Concatenate (SQL union) the files from step 10.
12. Aggregate the file from step 6 on variables FileInd and CohortInd. Set Count with the resulting counts.
13. Set DPID_SiteID to the value for the health plan.
14. Sort the file by FileInd and CohortInd, name it as ***DPID_SiteID*_NDI_Saved_Match_Results** and save to the MSOC data library.

IV. NDI Death Table Characterization

This file will be returned to the[Lead Analytic Site]It will contain aggregate counts of saved NDI death records, and will follow the data dictionary shown below.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
DeathYear	Num(3)	4.	Death year	2000-2018	Based on NDI date of death
State_Death_Code	Char(3)	$3.	State of death code	See Appendix A	
Count	Num(8)	comma10.	Count	0+	Count of records

Methods for creation of the NDI Death Table Characterization

1. Use the file **NDI_Death** from Section I, keeping variables DeathDt and State_Death_Code.
    a. Set DeathYear based on DeathDt.
2. Aggregate the file on variables DeathYear and State_Death_Code. Set Count with the resulting counts.
3. Set DPID_SiteID to the value for the health plan.
4. Sort the file by DeathYear and State_Death_Code, name it as ***DPID_SiteID*_NDI_Death_Char** and save to the MSOC data library.

V. NDI Cause of Death Table Characterization

This file will be returned to the [Lead Analytic Site]. It will contain aggregate counts of saved NDI cause of death records and will follow the data dictionary shown below.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CauseType	Char(3)	$3.	Cause type	C, I, O, R, U, 358, 113, 130	
CauseCat	Char(1)	$1.	Cause category	E, R, U	
Position1	Char(1)	$1.	Cause position	Part/line number on death certificate 1-6, R, U	
Position2	Num(3)	1.	Position Sequence	1-7 or missing	Sequence of condition within part/line code
Cause_Sequence	Num(3)	2.	Sequence Number	1-20 or missing	Up to 20 cause of death codes are available for each of entity COD and record COD
Nature_Injury	Char(1)	$1.	Nature of Injury Code	1, 0, N	
Count	Num(8)	comma10.	Count	0+	Count of records
Patients	Num(8)	comma10.	Patient count	0+	Count of patients in **NDI_COD** file

Methods for creation of the NDI Cause of Death Table Characterization

1. Use the file **NDI_COD** from Section II, keeping variables PatID, CauseType, CauseCat, Position1, Position2, Cause_Sequence, and Nature_Injury.
2. Aggregate the file on variables named in step 1, except for PatID. Set Count with the resulting counts.
3. Calculate the distinct number of PatIDs in the source file and save to variable Patients on each resulting row of the table. (This will be used for calculating proportions for selected stratifiers at the [list Lead analytic site]).
4. Set DPID_SiteID to the value for the health plan.
5. Sort the file by variables named in step 1, except for PatID, name it as ***DPID_SiteID*_NDI_COD_Char**, and save the file to the MSOC data library.

VI. Aggregate Descriptive Statistics by Cohort - Post Saved Matches

Using a version of the **All_Cohorts** file created in Package 1, this aggregate table will be returned to [list Lead analytic site].

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Dead	Num(3)	1.	Death flag	1=Dead 0=Alive	
Sex	Char(1)	$1.	Sex	M, F, missing	Missing for overall
AgeGroup	Char(9)	$9.	Age Group	Valid age groupings, missing	Missing for overall. Age groupings are: 45-54, 55-64, 65-74, 75-85, 86+ years
Torsades_Risk	Char(1)	$1.	Risk of Torsades de Pointes (TdP), per Credible Meds	K=Known P=Possible Risk C=Conditional Risk N=Not classified/unknown B=Baseline cohort default value	

Variable Name	Type & Length	Format	Label	Values	Definition / Comments / Guideline
Combined_Score	Char(5)	$5.	Comorbidity score category during baseline	Low-0 1 2-3 4-7 8+	
Year	Num(3)	4.	Index date year	2000+	
Variable	Char(30)	$30.	Variable	Age, NumVisits_IP, Num_Afib_flutter, etc.	Missing for sex, age group counts
Count	Num(8)	comma10.	Count	0+	
Percent	Num(8)	10.2	Percent	0-100	
Mean	Num(8)	10.2	Mean	0+	
StandardDeviation	Num(8)	10.2	Standard Deviation	0+	
Median	Num(8)	10.	Median	0+	
Min	Num(8)	10.	Min	0+	
Max	Num(8)	10.	Max	0+	

Methods to Create the Descriptive Aggregate file for [Lead Analytic Site]

1. Creating this table follows a similar process as in Package 1, Section IV, Aggregate Descriptive Statistics by Cohort, with the addition of the Dead variable (see step 4 below).
2. Check that the file named **All_Cohorts** is found in the DPLocal data library for Package 1.

   a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 1.
   b. Otherwise if found, create a temporary copy of the **All_Cohorts** file, saved in the DPLocal data library from package 1. Proceed with all other processes.
3. Determine if the health plan identifies a file that lists PatIDs excluded from research.
   a. If so, then on PatID, merge the temporary **All_Cohorts** file with such file, keeping all rows from temporary **All_Cohorts** containing PatIDs *not found* in the health plan file that lists PatIDs excluded from research.
   b. Otherwise, simply use the temporary **All_Cohorts** file and continue with next steps.
4. Create and add the Dead variable as follows:
   a. If the PatID from the **All_Cohorts** file is found in the **NDI_Death** file in this package, set as 1.
   b. If the PatID from the **All_Cohorts** file is not found in the **NDI_Death** file in this package, set as 0.
5. Using this temporary file, execute the following steps for each of the following 6 combinations of values of CohortInd and Dead: A "grouping" as used below is a combination of a value of CohortInd and a value of Dead.
   a. CohortInd=C1 and Dead=missing
   b. CohortInd=C1 and Dead=1
   c. CohortInd=C1 and Dead=0
   d. CohortInd=C2 and Dead=missing
   e. CohortInd=C2 and Dead=1
   f. CohortInd=C2 and Dead=0
6. Age Counts and Percents: Set Variable to "AgeGroup". Calculate the number of patients in each grouping and the percent that these counts represent. Collapse the patient-level AgeGroup to these groups of values: 45-54, 55-64, 65-74, 75-85, 86+ years. Calculate and fill rows for three groupings of Sex with these details:
   a. Overall: Use the count of all patients in each grouping as the denominator for calculating percents. Leave Sex blank, set AgeGroup to one of the collapsed groupings, Count and Percent to the respective calculated values. NOTE: In prior packages, there were no resulting rows with Sex blank and calculations across age groups; there were rows with Sex filled and AgeGroup blank, as well as both Sex and AgeGroup filled.
   b. Males and Females: Within each cohort and value of Dead, use the count of all patients in each grouping, with the respective sex, as the denominator for calculating percents. Set Sex=M or F appropriately, AgeGroup to one of the collapsed groupings, Count and Percent to the respective calculated values.
   c. Leave blank Torsades_Risk, Combined_Score, Year, Mean, StandardDeviation, Median, Min, and Max.

7. Age Descriptives: Set Variable to "Age". Calculate the following statistics: Mean, StandardDeviation, Median, Min, and Max. Calculate and fill rows for three groupings of Sex with these details:
    a. Overall: Calculate for all patients in each grouping. Leave blank Sex, AgeGroup, Torsades_Risk, Combined_Score, Year, Count and Percent. Fill Mean, StandardDeviation, Median, Min, and Max.
    b. Males and Females: Calculate for all patients in each grouping by sex. Leave blank AgeGroup, Torsades_Risk, Combined_Score, Year, Count and Percent. Fill Sex=M or F appropriately, Mean, StandardDeviation, Median, Min, and Max.
    c. Combined by Sex and AgeGroup: Calculate for all patients in each grouping by Sex and AgeGroup. Leave blank Torsades_Risk, Combined_Score, Year, Count and Percent. Fill Sex=M or F and AgeGroup appropriately, Mean, StandardDeviation, Median, Min, and Max. NOTE: In prior packages, there were descriptive rows for only Sex alone, Sex, combined with AgeGroup, but not for AgeGroup alone.
8. Sex Percent: Set Variable to "Sex" and Sex variable to "F" or "M" appropriately. Within each grouping, calculate the count and percent of patients that are female and fill both Count and Percent with these values. Leave blank AgeGroup, Torsades_Risk, Combined_Score, Year, Mean, StandardDeviation, Median, Min, and Max
9. Healthcare Utilization: Perform these processes for the following source variables in this order: NumVisits_IP, NumVisits_ED, NumVisits_AV, NumVisits_IS, NumVisits_OA, and RawDisp.
    a. Set Variable to the source variable named that is being used for calculations.
    b. Leave blank Sex, AgeGroup, Torsades_Risk, Combined_Score, Year, Count, and Percent.
    c. Using the source variable, calculate Mean, StandardDeviation, Median, Min, and Max; then fill these variables with the appropriate values.
10. Comorbid Conditions: Perform these processes for the following source variables, in this order: Num_Arrhythmia_Any, Num_Afib_Flutter, Num_PVT, Num_VF_Flutter, Num_PST, Num_Unsp_PT, Num_Prem_Beat, Num_Other_Dysrhythmia, Num_Arrhythmia_Any, Num_Afib_Flutter, Num_PVT, Num_VF_Flutter, Num_Cerebro, Num_CHD, Num_Diabetes, Num_Heart_Failure, Num_Defrib, Num_Hyperlipidemia, Num_Hypertension, Num_Kidney, Num_Circulatory, Num_Seizures, Num_Smoking, and Num_Obesity.
    a. Set Variable to the source variable named, that is being used for calculations.
    b. Leave blank Sex, AgeGroup, Torsades_Risk, Combined_Score, Year, Mean, StandardDeviation, Median, Min, and Max.
    c. Using the source variable, calculate Count and Percent, using the count of patients within a grouping as the denominator, and then fill these variables with the appropriate values.
11. Comorbidity Score Descriptives: Using the source variable Combined_Score_Num, perform the following:
    a. Set Variable to Combined_Score_Num.
    b. Leave blank Sex, AgeGroup, Torsades_Risk, Combined_Score, Year, Count, and Percent.

   c. Using the source variable, calculate Mean, StandardDeviation, Median, Min, and Max; then fill these variables with the appropriate values.

12. Torsades de pointes: Using the source variable Torsades_Risk, perform the following:
    a. Set Variable to Torsades_Risk.
    b. Leave blank Sex, AgeGroup, Combined_Score, Year, Mean, StandardDeviation, Median, Min, and Max.
    c. For each of these levels of Torsades_Risk: K, P, C, N, and B, calculate Count and Percent, using the count of patients within a grouping as the denominator, and then fill these variables with the appropriate values.
13. Comorbidity Score counts and percents: Using the source variable Combined_Score, perform the following:
    a. Set Variable to Combined_Score.
    b. Leave blank Sex, AgeGroup, Torsades_Risk, Year, Mean, StandardDeviation, Median, Min, and Max.
    c. For each of these levels of Combined_Score: Low-0, 1, 2-3, 4-7, 8+, calculate Count and Percent, using the count of patients within a cohort and value of Dead as the denominator, and then fill these variables with the appropriate values.
14. Cohort entry year: Using the source variable Year, perform the following:
    a. Set Variable to Year
    b. Leave blank Sex, AgeGroup, Torsades_Risk, Combined_Score, Mean, StandardDeviation, Median, Min, and Max.
    c. Using the source variable, calculate Count and Percent, using the count of patients within a cohort and value of Dead as the denominator, and then fill these variables with the appropriate values.
15. Set DPID_SiteID to the value for the health plan.
16. Name the resulting file as ***DPID_SITEID*_ndi_saved_descriptives_6** and save to the MSOC data library for this package.
17. Save the temporary **All_Cohorts** file as **All_Cohorts_6** to the DPLocal data library for this package, for use in a later step.

VII. Aggregate Comparison of Death Sources for Specific States

This file will be returned to the [Lead Analytic Site]. It enables comparison of death sources between health plans' death data from matching with their state death registry and NDI ascertainment of death. This is the dataset that will be returned to [Lead Analytic Site]. It will be used to create display table found in Appendix C. It will be restricted to NDI data for the state in which a health plan is located.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(3)	$3.	Cohort ID	C1=Exposed cohort C2=Baseline cohort ALL=Both cohorts	
Algorithm_Number	Num(3)	1.	Algorithm number	1, 2, or 3	Identifies the algorithm used to create matches from NDI records
Source	Char(1)	1.	Source of death information	L = Other, locally defined N = National Death Index S = State Death files T = Tumor data A = All	Each of these distinct; not crossed with Confidence
Confidence	Char(1)	1.	Confidence that the patient drawn from the Source data represents the actual patient (contrasts with Confidence in the Cause of Death table).	E = Excellent F = Fair P = Poor A = All	Each of these distinct; not crossed with Source

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Factor	Char(20)	$20.	Row categories	See Package 5 specifications, Appendix C	Factors that define rows
True_Pos	Num(8)	comma10.	Count (A)	0+	Count of death in health plan and death in NDI
False_Pos	Num(8)	comma10.	Count (B)	0+	Count of alive in health plan and death in NDI
False_Neg	Num(8)	comma10.	Count (C)	0+	Count of death in health plan and alive in NDI
True_Neg	Num(8)	comma10.	Count (D)	0+	Count of alive in health plan and alive in NDI
Test_Positive	Num(8)	comma10.	Count (A +B)	0+	Count of True_Pos (A) + False_Pos (B)
Test_Negative	Num(8)	comma10.	Count (C +D)	0+	Count of False_Neg (C) + True_Neg (D)
Plan_Death	Num(8)	comma10.	Count (A + C)	0+	Count of True_Pos (A) + False_Neg (C)
Plan_NoDeath	Num(8)	comma10.	Count (B +D)	0+	Count of False_Pos (B) + True_Neg (D)
Total_Pop	Num(8)	comma10.	Count of population	0+	Count of True_Pos (A) + False_Pos (B) + False_Neg (C) + True_Neg (D)
Sensitivity	Num(8)	5.2	Sensitivity	000.00% +	True_Pos (A) divided by [(True_Pos (A) + False_Neg (C)]
Specificity	Num(8)	5.2	Specificity	000.00% +	True_Neg (D) divided by [False_Neg (C) + True_Neg (D)]
PPV	Num(8)	5.2	Positive Predictive Value	000.00% +	True_Pos (A) divided by [True_Pos (A) + False_Pos (B)]

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
NPV	Num(8)	5.2	Negative Predictive Value	000.00% +	True_Neg (D) divided by [False_Neg (C) + True_Neg (D)]

Methods for creation of aggregate of matching to Data Partner death data

1. Using the table referenced by Appendix A, identify the row with a matching DPID_SiteID to that of the health plan executing the package.
    a. If DeathTable=Y, then save the value for State to temporary variable State; continue with step 2.
    b. If DeathTable=N, do not continue processing in this section as the DPID_SiteID does not have a single state in which it matches patient records.
2. Check that the file named **NDI_Outbound_Records** is found in the DPLocal data library for Package 4.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal data library from Package 4.
    b. Otherwise if found, proceed with all other processes.
3. Use the patient-level file **NDI_Selection** from Package 5. (Determining availability of this file will have already occurred in Section I Include from this file only those PatIDs that are found in the **All_Cohorts_6** file that was created in Section VI. (This will maintain analysis on only those patients who are approved for research projects.) Then filter this file for only those rows where State_Death_Code = the State saved in step 1.a.
4. From Package 5, v2.3.3, Section VIII (Aggregate Comparison of Death Sources), run through steps 3 to 14.
5. Set DPID_SiteID to the value for the health plan.
6. Sort the table by CohortInd, Algorithm, and Factor. Name the table ***DPID_SiteID*_NDI_Death_Compare_6** and save to the MSOC data library.

VIII. Aggregate Incidence of Cardiovascular Death (A2)

References within parentheses are table numbers from the protocol (e.g., A2). This file will be returned to the [Lead Analytic Site]. It will contain aggregates of selected NDI death and cause data along with cohort data and will follow the data dictionary shown below. It will be used to create the display table found in Appendix D. This is one example of a table that can be produced for use in future analyses, but multiple similar tables were created for the pilot project.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
Group	Char(25)	25.	Patient group	See Appendix E	
CVD_C1_Inc	Num(8)	5.2	CVD Incidence Antiarrhythmic medication users in cohort C1	0-1	
CVD_C2_Inc	Num(8)	5.2	CVD Incidence in cohort C2	0-1	
Dth_C1_Inc	Num(8)	5.2	Death Incidence Antiarrhythmic medication users in cohort C1	0-1	
Dth_C2_Inc	Num(8)	5.2	Death Incidence in cohort C2	0-1	
CVD_C1_Ct	Num(8)	comma10.	Count of CVD death in cohort C1	0+	
CVD_C2_Ct	Num(8)	comma10.	Count of CVD death in cohort C2	0+	
Dth_C1_Ct	Num(8)	comma10.	Count of all cause death in cohort C1	0+	
Dth_C2_Ct	Num(8)	comma10.	Count of all cause death in cohort C2	0+	
C1_Ct	Num(8)	comma10.	Count of all patients in cohort C1	0+	
C2_Ct	Num(8)	comma10.	Count of all patients in cohort C2	0+	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
C1_COD_Ct	Num(8)	comma10.	C1 denominator for CVD	0+	C1 patients with either: - No NDI death - NDI death and COD records
C2_COD_Ct	Num(8)	comma10.	C2 denominator for CVD	0+	C2 patients with either: - No NDI death - NDI death and COD records

Methods for creation of aggregate of matching to Data Partner death data

1. Use the **All_Cohorts_6** table from Section VI. Include variables PatID, CohortInd, Sex, Age, Num_CHD, and Num_Diabetes. By PatID, link with **NDI_Death** from Section I. Create temporary variable Death and set as follows:
    a. If a PatID from **All_Cohorts_6** exists in **NDI_Death**, set to 1.
    b. If a PatID from **All_Cohorts_6** does not exist in **NDI_Death**, set to 0.
2. By PatID, link the file from step 1 with file **NDI_COD** from Section II. , only for rows where CauseType=U. Keep variable COD for these rows. Create temporary variable COD_Available and set as follows:
    a. If Death=1, but no rows from **NDI_COD** where CauseType=U contain the PatID from **NDI_Death**, then set COD_Available =0. This is also the case if a PatID from step 1 is found in **NDI_COD** but no rows have CauseType=U.
    b. If Death=1 and a row from **NDI_COD** where CauseType=U contains the PatID from **NDI_Death**, then set COD_Available =1.
    c. If Death=0, set to missing.
3. Set additional temporary patient level variables as follows:
    a. CVD_C1: For patients with CohortInd=C1, set as 1 if COD from step 2 is one of the codes for CVD (Sample codes only are found in Appendix F). Set as 0 if either COD is not one of the codes found in Appendix F or Death=0.
    b. CVD_C2: For patients with CohortInd=C2, set as 1 if COD from step 2 is one of the codes for CVD (Sample codes only are found in Appendix F). Set as 0 if either COD is not one of the codes found in Appendix F or Death=0.
    c. Dth_C1: For patients with CohortInd=C1, set equal to Death value from step 1.

   d. Dth_C2: For patients with CohortInd=C2, set equal to Death value from step 1.

4. Create aggregates for each Group defined in Appendix E, following the definition indicated:
   a. Set the Group variable, concatenating Appendix E Sort_Order, a hyphen, and the Appendix E Group value.
   b. C1_Ct: Count of all patients with CohortInd=C1.
   c. C2_Ct: Count of all patients with CohortInd=C2.
   d. CVD_C1_Ct: Sum of CVD_C1 from step 3.a.
   e. CVD_C2_Ct: Sum of CVD_C2 from step 3.b.
   f. Dth_C1_Ct: Sum of Dth_C1 from step 3.b.
   g. Dth_C2_Ct: Sum of Dth_C2 from step 3.d.
   h. Dth_C1_Inc: Dth_C1_Ct divided by C1_Ct, multiplied by 100.
   i. Dth_C2_Inc: Dth_C2_Ct divided by C2_Ct. multiplied by 100.
   j. C1_COD_Ct: C1_Ct minus count of C1 patients with COD_Available=0.
   k. C2_COD_Ct: C2_Ct minus count of C2 patients with COD_Available=0.
   l. CVD_C1_Inc: CVD_C1_Ct divided by C1_COD_Ct, multiplied by 100.
   m. CVD_C2_Inc: CVD_C2_Ct divided by C2_COD_Ct, multiplied by 100.
5. Set DPID_SiteID to the value for the health plan.
6. Fill the variables in order listed and as per the data dictionary above.
7. Sort the file by Sort_Order and save as ***DPID_SiteID*_CVD_Aggregate_A2** in the MSOC data library.

IX. Outbound Records to NDI With Selection Reason

This is a patient-level file of *records* submitted to the NDI from Package 4. It accounts for records eligible for research as of both Packages 5 and 6.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char(#)	$#.	Patient ID	Alpha-Numeric	Arbitrary person-level identifier. Used to link across tables.
PackageInd	Char(1)	$1.	Package indicator	5=As of Package 5 6=As of Package 6	Indicator of data source package
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	
Control_Number	Char(10)	$10.	Control number	0000000001+	Sequential number, with leading zeros, incremented for each PatID written to this table, inserted automatically
Sequence_Number	Char(2)	$2.	Sequence number	01+	Sequential number, with leading zeros, incremented for each separate row, per PatID, written to this table, inserted automatically. Each PatID will have at least one row with a value of "01"
Reason_Select	Char(3)	$3.	Reason record selected	DE=Death DI=Disenrolled B=Both LTF=Lost to follow-up	Reason selected: death, disenrollment, both, or lost to follow-up

Methods to Create the Outbound Records to NDI With Selection Reason File

1. Set PackageInd=5. Check that the files named **NDI_Selection, Potential_Ties**, and **Duplicates** are found in the DPLocal data library from Package 5.
    a. If any of the three files are not found, abort this process and print an error message to the log indicating the name of the file(s) that must be in the DPLocal library.
    b. Otherwise if found, proceed with all other processes.
2. Concatenate all three files, keeping variables PatID, CohortInd, Control_Number, Sequence_Number, and Reason_Select.
3. Resolve to unique rows on the basis of the variables named in step 2, to a temporary file.
4. Set PackageInd=6. Check that the file named **NDI_Outbound_Records** is found in the DPLocal data library from Package 4.
    a. If not found, abort this process and print an error message to the log indicating that this file must be in the DPLocal library.
    b. Otherwise if found, proceed with all other processes.
5. From the **NDI_Outbound_Records** file, keep variables PatID, CohortInd, Control_Number, Sequence_Number.
6. Link on PatID with the **All_Cohorts_6** file from Section VI to obtain variables DeathSrc, and FinalEnrollmentStatus for the next step, keeping all rows from **NDI_Outbound_Records** only.
7. Create variable Reason_Select, using variables DeathSrc and FinalEnrollmentStatus as follows:
    a. DE = DeathSrc equal to D, E, or B
    b. DI = FinalEnrollmentStatus equal to D
    c. B = Both steps 7.a and 7.b are true
    d. LTF (for “lost to follow-up” = Neither steps 7.a nor 7.b are true)
8. Concatenate (i.e., SQL UNION) the files resulting from steps 3 and 7.
9. Save the file as **NDI_Outbound_Reason_Select** to the DPLocal Data Library.

X. Selection Reason Aggregate

This is an aggregate file returned to the[Lead Analytic Site]It is used for characterization and for establishing denominators for other aggregate files.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
PackageInd	Char(1)	$1.	Package indicator	5=As of Package 5 6=As of Package 6	Indicator of data source package
CohortInd	Char(3)	$3.	Cohort indicator	C1 = Antiarrhythmic cohort C2= Baseline cohort	CohortInd
Reason_Select	Char(3)	$3.	Reason record selected	DE=Death DI=Disenrolled B=Both LTF=Lost to followup	Reason selected: death, disenrollment, both, or lost to follow-up
Patients_Ct	Num(8)	comma10.	Distinct patients	0+	Count of patients across stratifiers
Records_Ct	Num(8)	comma10.	Distinct records	0+	Count of records across stratifiers

Methods to Create the Selection Reason Aggregate File

1. Use the **NDI_Outbound_Reason_Select** file created in Section IX.
2. Aggregate the entire file on PackageInd, CohortInd and Reason_Select, setting Records_Ct to the resulting counts.
3. Create a temporary file of distinct PatID, PackageInd, CohortInd, and Reason_Select.
    a. Aggregate this temporary file on PackageInd, CohortInd and Reason_Select, setting Patients_Ct to the resulting counts.
4. Set DPID_SiteID to the value for the health plan for the variable in the table and the table name.
5. Sort the file by CohortInd and Reason_Select. Save to the MSOC data library as ***DPID_SiteID*_Select_Reason_Agg**.

XI. Death Causes File

This is a patient-level file, with indicators of patients meeting various criteria. It serve as numerators and denominators for selected aggregate files. The patient-level file remains at the Health Plan.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
PatID	Char(#)	Site specific length	Patient ID	Alpha-Numeric	#= Site specific length
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	Package 1 All_Cohorts file
Sex	Char(1)	$1.	Sex of patient	M, F only	Package 1 All_Cohorts file
Age	Num(3)	3.	Age at index date	45+	Package 1 All_Cohorts file
Birth_Date	Num(4)	mmddyy10.	Date of birth	Valid SAS dates	Package 1 All_Cohorts file
IndexDt	Num(4)	mmddyy10.	Index date of drug therapy	Valid SAS dates	Package 1 All_Cohorts file
IndexYear	Num(3)	4.	Index year of drug therapy	2000-2018	Calculated
EpisodeEndDt	Num(4)	mmddyy10.	Exposure episode end date	Valid SAS dates	Package 1 All_Cohorts file
Episode365Dt	Num(4)	mmddyy10.	Exposure episode end date + 365 days	Valid SAS dates	Calculated
Death	Num(3)	1.	NDI death	1=Yes 0=No	NDI_Death
DeathDt	Num(4)	mmddyy10.	NDI Date of death	Valid SAS date or missing	As appears on the NDI record
DeathYear	Num(3)	4.	Index year of drug therapy	2000-2018 or missing	Calculated

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
COD	Char(8)	$8.	NDI underlying cause of death	Free text or missing	NDI_Cause: All are assuming to be ICD10
COD_Link	Num(3)	1.	NDI COD records linked	1=Yes 0=No Missing	NDI_Cause
COD_U_Available	Num(3)	1.	NDI underlying COD record available	1=Yes 0=No Missing	NDI_Cause
COD_U_Multiple	Num(3)	1.	More than 1 underlying COD record available	1=Yes 0=No Missing	NDI_Cause
Death_Cause_Denom	Num(3)	1.	Mortality cause denominator flag	1=Yes 0=No	All_Cohort, NDI_Death, and NDI_COD
CVD	Num(3)	1.	NDI COD for cardiovascular death	1=CVD 0=Not CVD	

Methods for creation of patient level Death Causes file

1. Use the **All_Cohorts_6** table from Section VI. Include variables PatID, CohortInd, Sex, Age, Birth_Date, IndexDt, and EpisodeEndDt.
    a. Set Episode365Dt as EpisodeEndDt + 365.
    b. Set IndexYear as year part of IndexDt.

2. By PatID, link with **NDI_Death** from Section II, keeping all rows from **All_Cohorts_6**. Create temporary variable Death and set as follows:
    a. If a PatID from **All_Cohorts_6** exists in **NDI_Death**, set Death to 1.
        i. Keep variable DeathDt.
        ii. Set DeathYear as year part of DeathDt.
    b. If a PatID from **All_Cohorts_6** does not exist in **NDI_Death**, set Death to 0.
        i. Set DeathDt and DeathYear to missing.
3. By PatID, link the file from step 2 with file **NDI_COD** from Section II, keeping all rows from the step 2 file.
    a. Create variable COD_Link and set as follows:
        i. If Death=1 and a row from **NDI_COD** contains the PatID from **NDI_Death**, then set COD_Link=1.
        ii. If Death=1 but no row from **NDI_COD** exists where then set COD_Link=0.
        iii. If Death=0, set to missing.
    b. Create variable COD_U_Available and set as follows:
        i. If Death=1 and a row from **NDI_COD** where CauseType=U contains the PatID from **NDI_Death**, then set COD_U_Available =1. Keep variable COD only if this condition is true. If there is more than one row per PatID, where CauseType=U, perform the following:
            1. Create distinct rows for each COD value where **NDI_COD** Cause_Type=U. All other variables should have the same values per PatID.
            2. Set COD_U_Multiple=1.
        ii. If there is only 1 **NDI_COD** row with Cause_Type=U, set COD_U_Multiple=0.
        iii. If Death=1 but no row from **NDI_COD** exists where CauseType=U, and contains the same PatID from **NDI_Death**, then set COD_U_Available =0 and COD_U_Multiple=missing. This is also the case if a PatID from **NDI_Death** is found in **NDI_COD** but no rows have CauseType=U. Do not keep variable COD for these rows if this condition is true.
        iv. If Death=0, set COD_U_Available= missing and COD_U_Multiple=missing.
4. Death_Cause_Denom flag variable. This includes patients who are alive or for whom both a **NDI_Death** record and a **NDI_COD** exist:
    a. Set as 1 if Death=0, or both Death=1 and COD_U_Available=1.
    b. Otherwise, set to 0. This will be only for those patients who had Death=1 but COD_U_Available=0.
5. Cause of death flags:
    a. CVD: Set as 1 if any COD per PatID is one of the codes for CVD (Sample codes only are found in Appendix F). Set as 0 if either COD is not one of the codes found in Appendix F or Death=0.
6. Sort by PatID and CohortInd. Save to the DPLocal data library as **Death_Causes_File**.

XII. Aggregate COD Completeness File

This file will be returned to the [Lead Analytic Site]. It will contain aggregates of whether an NDI death record also has cause data. This step was implemented because the NDI states that cause of death information is not always returned for all records. It will further indicate, if cause data is available, whether there is always an underlying cause of death record or not, as well as if there is more than one underlying cause of death record.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
CohortInd	Char(2)	$2.	Cohort indicator	C1 = Antiarrhythmic cohort C2 = Baseline cohort	
Death	Num(3)	1.	NDI death	1=Yes 0=No	
COD_Link	Num(3)	1.	NDI COD records linked	1=Yes 0=No Missing	
COD_U_Available	Num(3)	1.	NDI underlying COD record available	1=Yes 0=No Missing	
COD_U_Multiple	Num(3)	1.	More than 1 underlying COD record available	1=Yes 0=No Missing	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Records_Ct	Num(8)	comma10.	Distinct records	0+	Count of records across stratifiers

Methods for creation of patient level COD Completeness file

1. Use the **Death_Causes_File** table from Section IX. Include variables PatID, CohortInd, Death, COD_Link, COD_U_Available and COD_U_Multiple. Resolve to a single row per PatID of all of these variables.
2. Aggregate the file on the variables in step 1, except for PatID, setting Records_Ct to the counts per stratification.
3. Fill the file according to the data dictionary above, preserving the variable order.
4. Set DPID_SiteID to the value for the health plan for the variable in the table and the table name.
5. Sort by CohortInd and save to the MSOC data library as file ***DPID_SiteID*_COD_Completeness**

XIII. XIII. Aggregate Death Causes File (A3)

This file will be returned to the Harvard Pilgrim Health Care Institute. It will contain aggregates of selected NDI death and cause data along with cohort data, and will follow the data dictionary shown below. It will be used to create the display table found in Appendix G. Description of Type 1 and Type 2 denominators and numerators:

Form 1: We will follow each patient for death from cohort entry through the end of NDI data; end of NDI data varies by plan. Follow up will be truncated at death, or at end of NDI data, whichever comes first.

Form 2: We will follow each patient for death from cohort entry through end of episode plus 365 days (mirrors our NDI submission criteria). Only deaths that occur during follow-up will be counted. Follow up will be truncated at death, or at end of episode plus 365 days, whichever comes first.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
Sort_Order	Num(3)	3.	Sort Order	1-218; see Appendix H	
Year	Num(3)	4.	Year	2000-2018	
Group	Char(30)	30.	Patient group	See Appendix H	
CVD_Form1_C1_Inc	Num(8)	5.2	CVD Incidence Antiarrhythmic medication users in cohort C1, Form 1	0-1	
CVD_Form1_C2_Inc	Num(8)	5.2	CVD Incidence in cohort C2, Form 1	0-1	
CVD_Form2_C1_Inc	Num(8)	5.2	CVD Incidence Antiarrhythmic medication users in cohort C1, Form 2	0-1	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
CVD_Form2_C2_Inc	Num(8)	5.2	CVD Incidence in cohort C2, Form 2	0-1	
Dth_Form1_C1_Inc	Num(8)	5.2	Death Incidence Antiarrhythmic medication users in cohort C1, Form 1	0-1	
Dth_Form1_C2_Inc	Num(8)	5.2	Death Incidence in cohort C2, Form 1	0-1	
Dth_Form2_C1_Inc	Num(8)	5.2	Death Incidence Antiarrhythmic medication users in cohort C1, Form 2	0-1	
Dth_Form2_C2_Inc	Num(8)	5.2	Death Incidence in cohort C2, Form 2	0-1	
CVD_Form1_C1_Ct	Num(8)	comma10.	Count of CVD Form 1 death in cohort C1	0+	
CVD_Form1_C2_Ct	Num(8)	comma10.	Count of CVD Form 1 death in cohort C2	0+	
CVD_Form2_C1_Ct	Num(8)	comma10.	Count of CVD Form 2 death in cohort C1	0+	
CVD_Form2_C2_Ct	Num(8)	comma10.	Count of CVD Form 2 death in cohort C2	0+	
Dth_Form1_C1_Ct	Num(8)	comma10.	Count of all cause Form 1 death in cohort C1	0+	
Dth_Form1_C2_Ct	Num(8)	comma10.	Count of all cause Form 1 death in cohort C2	0+	
Dth_Form2_C1_Ct	Num(8)	comma10.	Count of all cause Form 2 death in cohort C1	0+	
Dth_Form2_C2_Ct	Num(8)	comma10.	Count of all cause Form 2 death in cohort C2	0+	
Dth_Denom_Form1_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, Form 1 denominator for death	0+	
Dth_Denom_Form1_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, Form 1 denominator for death	0+	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Dth_Denom_Form2_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, Form 2 denominator for death	0+	
Dth_Denom_Form2_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, Form 2 denominator for death	0+	
Cause_Denom_Form1_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, Form 1 denominator for cause of death	0+	
Cause_Denom_Form1_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, Form 1 denominator for cause of death	0+	
Cause_Denom_Form2_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, Form 2 denominator for cause of death	0+	
Cause_Denom_Form2_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, Form 2 denominator for cause of death	0+	

Methods for creation of patient level Death Causes aggregate file A3

1. Use the **Death_Causes_File** file from Section XI, keeping variables: PatID, CohortInd, Sex, Birth_Date, IndexDt, IndexYear, Episode365Dt, Death, DeathYear, Death_Cause_Denom, and CVD. Resolve to a single row per PatID of all of these variables, setting CVD to the maximum value found per PatID.
2. Using Appendix H, create successive subsets of patients. Set Sort_Order, Year, and Group as defined in the appendix. Then follow all steps for each subset, filling variables as defined. Note that Age has to be recomputed as of January 1 of each calendar year, using the Birth_Date variable from the **Death_Causes_File**; also note Appendix H instructions for females at age 54 as of IndexDt and for males at age 44 as of IndexDt. When the instructions below say:
    a. IndexDt through Episode365Dt overlaps Year: an overlap is defined as 1 or more days of overlap of the interval and the calendar year of calculation.

   b. DeathYear > Year: Patients are counted *within* the Year in which DeathDt occurs and earlier; patient are *not* counted in subsequent years.
   c. Either one of the following is true: DeathYear is missing or DeathYear >= Year: Either can be true as a logical "OR" condition.

3. Dth_Denom_Form1_C1_Ct: Count of patients where CohortInd="C1" and IndexYear <= Year and either one of the following is true: DeathYear is missing or DeathYear >= Year.
4. Dth_Denom_Form1_C2_Ct: Count of patients where CohortInd="C2" and IndexYear <= Year and either DeathYear is missing or DeathYear >= Year.
5. Dth_Denom_Form2_C1_Ct: Count of patients where CohortInd="C1" and IndexDt through Episode365Dt overlaps Year and either DeathYear is missing or DeathYear >= Year.
6. Dth_Denom_Form2_C2_Ct: Count of patients where CohortInd="C2" and IndexDt through Episode365Dt overlaps Year and either DeathYear is missing or DeathYear >= Year.
7. Cause_Denom_Form1_C1_Ct: Count of patients where CohortInd="C1", IndexYear <= Year and Death_Cause_Denom=1.
8. Cause_Denom_Form1_C2_Ct: Count of patients where CohortInd="C2", IndexYear <= Year and Death_Cause_Denom=1.
9. Cause_Denom_Form2_C1_Ct: Count of patients where CohortInd="C1", IndexDt through Episode365Dt overlaps Year and Death_Cause_Denom=1.
10. Cause_Denom_Form2_C2_Ct: Count of patients where CohortInd="C2", IndexDt through Episode365Dt overlaps Year and Death_Cause_Denom=1.
11. Dth_Form1_C1_Ct: Count of patients where CohortInd="C1", year of DeathYear = Year and Death=1.
12. Dth_Form1_C2_Ct: Count of patients where CohortInd="C2", year of DeathYear = Year and Death=1.
13. Dth_Form2_C1_Ct: Count of patients where CohortInd="C1", IndexDt through Episode365Dt overlaps Year, DeathYear = Year, and Death=1.
14. Dth_Form2_C2_Ct: Count of patients where CohortInd="C2", IndexDt through Episode365Dt overlaps Year, DeathYear = Year, and Death=1.
15. CVD_Form1_C1_Ct: Count of patients where CohortInd="C1", DeathYear = Year, Death=1, and CVD=1.
16. CVD_Form1_C2_Ct: Count of patients where CohortInd="C2", DeathYear = Year, Death=1, and CVD=1.
17. CVD_Form2_C1_Ct: Count of patients where CohortInd="C1", IndexDt through Episode365Dt overlaps Year, DeathYear = Year, Death=1, and CVD=1.
18. CVD_Form2_C2_Ct: Count of patients where CohortInd="C2", IndexDt through Episode365Dt overlaps Year, DeathYear = Year, Death=1, and CVD=1.
19. Dth_Form1_C1_Inc: Dth_Form1_C1_Ct divided by Dth_Denom_Form1_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
20. Dth_Form1_C2_Inc: Dth_Form1_C2_Ct divided by Dth_Denom_Form1_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
21. Dth_Form2_C1_Inc: Dth_Form2_C1_Ct divided by Dth_Denom_Form2_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
22. Dth_Form2_C2_Inc: Dth_Form2_C2_Ct divided by Dth_Denom_Form2_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
23. CVD_Form1_C1_Inc: CVD_Form1_C1_Ct divided by Cause_Denom_Form1_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
24. CVD_Form1_C2_Inc: CVD_Form1_C2_Ct divided by Cause_Denom_Form1_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
25. CVD_Form2_C1_Inc: CVD_Form2_C1_Ct divided by Cause_Denom_Form2_C1_Ct, multiplied by 100 and rounded to 2 decimal places.

26. CVD_Form2_C2_Inc: CVD_Form2_C2_Ct divided by Cause_Denom_Form2_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
27. Fill the file according to the data dictionary above, preserving the variable order. Sort by Sort_Order and save to the MSOC data library as file ***DPID_SiteID*_Death_Aggregate_A3**.

XIV. Aggregate Death Causes File (A4)

This file will be returned to the Harvard Pilgrim Health Care Institute. It will contain aggregates of selected NDI death and cause data along with cohort data, and will follow the data dictionary shown below. It will be used to create the display table found in Appendix I. Description of Form 1 and Form 2 denominators and numerators:

Form 1: We will follow each patient for death from cohort entry through the end of NDI data; end of NDI data varies by plan. Follow up will be truncated at death, or at end of NDI data, whichever comes first.

Form 2: We will follow each patient for death from cohort entry episode end plus through 365 days following episode end (mirrors our NDI submission criteria). Only deaths that occur during follow-up will be counted. Follow up will be truncated at death, or at end of episode plus 365 days, whichever comes first.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
Years	Char(9)	$9.	Years	2000-2004, 2005-2009, 2010-2014, 2015-2017, 2018	Each year grouping is distinct
CVD_Form1_C1_Inc	Num(8)	5.2	CVD Incidence Antiarrhythmic medication users in cohort C1, Type 1	0-1	
CVD_Form1_C2_Inc	Num(8)	5.2	CVD Incidence in cohort C2, Type 1	0-1	
CVD_Form2_C1_Inc	Num(8)	5.2	CVD Incidence Antiarrhythmic medication users in cohort C1, Type 2	0-1	
CVD_Form2_C2_Inc	Num(8)	5.2	CVD Incidence in cohort C2, Type 2	0-1	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Dth_Form1_C1_Inc	Num(8)	5.2	Death Incidence Antiarrhythmic medication users in cohort C1, Type 1	0-1	
Dth_Form1_C2_Inc	Num(8)	5.2	Death Incidence in cohort C2, Type 1	0-1	
Dth_Form2_C1_Inc	Num(8)	5.2	Death Incidence Antiarrhythmic medication users in cohort C1, Type 2	0-1	
Dth_Form2_C2_Inc	Num(8)	5.2	Death Incidence in cohort C2, Type 2	0-1	
CVD_Form1_C1_Ct	Num(8)	comma10.	Count of CVD type 1 death in cohort C1	0+	
CVD_Form1_C2_Ct	Num(8)	comma10.	Count of CVD type 1 death in cohort C2	0+	
CVD_Form2_C1_Ct	Num(8)	comma10.	Count of CVD type 2 death in cohort C1	0+	
CVD_Form2_C2_Ct	Num(8)	comma10.	Count of CVD type 2 death in cohort C2	0+	
Dth_Form1_C1_Ct	Num(8)	comma10.	Count of all cause type 1 death in cohort C1	0+	
Dth_Form1_C2_Ct	Num(8)	comma10.	Count of all cause type 1 death in cohort C2	0+	
Dth_Form2_C1_Ct	Num(8)	comma10.	Count of all cause type 2 death in cohort C1	0+	
Dth_Form2_C2_Ct	Num(8)	comma10.	Count of all cause type 2 death in cohort C2	0+	
Cause_Denom_Form1_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, type 1 denominator for cause of death	0+	
Cause_Denom_Form1_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, type 1 denominator for cause of death	0+	
Cause_Denom_Form2_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, type 2 denominator for cause of death	0+	

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
Cause_Denom_Form2_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, type 2 denominator for cause of death	0+	
Dth_Denom_Form1_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, type 1 denominator for death	0+	
Dth_Denom_Form1_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, type 1 denominator for death	0+	
Dth_Denom_Form2_C1_Ct	Num(8)	comma10.	Count of patients in cohort C1, type 2 denominator for death	0+	
Dth_Denom_Form2_C2_Ct	Num(8)	comma10.	Count of patients in cohort C2, type 2 denominator for death	0+	

Methods for creation of patient level Death Causes aggregate file A4

1. Use the ***DPID_SiteID*_Death_Aggregate_A3** file from Section XIII, including all variables.
2. Perform the following aggregations across each value of Years (i.e., 2000-2004, 2005-2009, 2010-2014, 2015-2017, 2018). For example for Years=2000-2004, use the values when Sort_Order=1, 13, 25, 37, or 49, Year=2000 through 2004, and GroupName is one of: Year-2000, Year-2001, Year-2002, Year-2003, or Year-2004.
    a. Set Years= year grouping for each of the Valid Values in the dictionary.
    b. Dth_Denom_Form1_C1_Ct: Sum of Dth_Denom_Form1_C1_Ct.
    c. Dth_Denom_Form1_C2_Ct: Sum of Dth_Denom_Form1_C2_Ct.
    d. Dth_Denom_Form2_C1_Ct: Sum of Dth_Denom_Form2_C1_Ct.
    e. Dth_Denom_Form2_C2_Ct: Sum of Dth_Denom_Form2_C2_Ct.
    f. Cause_Denom_Form1_C1_Ct: Sum of Cause_Denom_Form1_C1_Ct.
    g. Cause_Denom_Form1_C2_Ct: Sum of Cause_Denom_Form1_C2_Ct.
    h. Cause_Denom_Form2_C1_Ct: Sum of Cause_Denom_Form2_C1_Ct.
    i. Cause_Denom_Form2_C2_Ct: Sum of Cause_Denom_Form2_C1_Ct.

j. Dth_Form1_C1_Ct: Sum of Dth_Form1_C1_Ct.
k. Dth_Form1_C2_Ct: Sum of Dth_Form1_C2_Ct.
l. Dth_Form2_C1_Ct: Sum of Dth_Form2_C1_Ct.
m. Dth_Form2_C2_Ct: Sum of Dth_Form2_C2_Ct.
n. CVD_Form1_C1_Ct: Sum of CVD_Form1_C1_Ct.
o. CVD_Form1_C2_Ct: Sum of CVD_Form1_C2_Ct.
p. CVD_Form2_C1_Ct: Sum of CVD_Form2_C1_Ct.
q. CVD_Form2_C2_Ct: Sum of CVD_Form2_C2_Ct.

3. Use the variables calculated in step 2 for the following:
   a. CVD_Form1_C1_Inc: CVD_Form1_C1_Ct divided by Cause_Denom_Form1_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
   b. CVD_Form1_C2_Inc: CVD_Form1_C2_Ct divided by Cause_Denom_Form1_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
   c. CVD_Form2_C1_Inc: CVD_Form2_C1_Ct divided by Cause_Denom_Form2_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
   d. CVD_Form2_C2_Inc: CVD_Form2_C2_Ct divided by Cause_Denom_Form2_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
   e. Dth_Form1_C1_Inc: Dth_Form1_C1_Ct divided by Dth_Denom_Form1_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
   f. Dth_Form1_C2_Inc: Dth_Form1_C2_Ct divided by Dth_Denom_Form1_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
   g. Dth_Form2_C1_Inc: Dth_Form2_C1_Ct divided by Dth_Denom_Form2_C1_Ct, multiplied by 100 and rounded to 2 decimal places.
   h. Dth_Form2_C2_Inc: Dth_Form2_C2_Ct divided by Dth_Denom_Form2_C2_Ct, multiplied by 100 and rounded to 2 decimal places.
4. Fill the file according to the data dictionary above, preserving the variable order.
5. Set DPID_SiteID to the value for the health plan.
6. Sort by Years and save to the MSOC data library as file ***DPID_SiteID*_Death_Aggregate_A4**.

.

Appendix A – State Postal Codes

State	Alpha Code	Numeric Code	State	Alpha Code	Numeric Code
Alabama	AL	01	New Jersey	NJ	31
Alaska	AK	02	New Mexico	NM	32
Arizona	AZ	03	New York	NY	33
Arkansas	AR	04	New York City	NYC	33C
California	CA	05	North Carolina	NC	34
Colorado	CO	06	North Dakota	ND	35
Connecticut	CT	07	Ohio	OH	36
Delaware	DE	08	Oklahoma	OK	37
District of Columbia	DC	09	Oregon	OR	38
Florida	FL	10	Pennsylvania	PA	39
Georgia	GA	11	Rhode Island	RI	40
Hawaii	HI	12	South Carolina	SC	41
Idaho	ID	13	South Dakota	SD	42
Illinois	IL	14	Tennessee	TN	43
Indiana	IN	15	Texas	TX	44
Iowa	IA	16	Utah	UT	45
Kansas	KS	17	Vermont	VT	46
Kentucky	KY	18	Virginia	VA	47
Louisiana	LA	19	Washington	WA	48
Maine	ME	20	West Virginia	WV	49
Maryland	MD	21	Wisconsin	WI	50
Massachusetts	MA	22	Wyoming	WY	51
Michigan	MI	23	Puerto Rico	PR	52
Minnesota	MN	24	Virgin Islands	VI	53
Mississippi	MS	25	Guam	GU	54
Missouri	MO	26	Canada	CN	55
Montana	MT	27	Cuba	CU	56
Nebraska	NE	28	Mexico	MX	57
Nevada	NV	29	Remainder of World	RW	59
New Hampshire	NH	30	Unknown		99

Appendix B – Health Plan States

[Lead Analytic Site] will create this input file.

Variable	Type (Length)	Format	Label	Valid Values	Source / Comments
DPID_SiteID	Char(6)	$6.	Site identifier	Free text	Concatenation of Sentinel DPID and SiteID
DeathTable	Char(1)	$1.	Death table available	Y=Yes, has a Death table N=No, has no Death table	DeathTable
State	Char(3)	$3.	State abbreviation	AL, AK, AZ, etc.	Postal code for state in which health plan is located
LastDeathDate	Num(4)	mmddyy10.	Final death date within truncation	Valid SAS dates	
NDI_Final_Year	Num(3)	4.	Final NDI year matched	2017-2018	Indicates last year with which the plan matched at NDI

DPID_SiteID	DeathTable	State	LastDeathDate	NDI_Final_Year
Health Plan 1	N	XX	x/xx/xxxx	xxxx
Health Plan 2	Y	XX	x/xx/xxxx	xxxx
Health Plan 3	Y	XX	x/xx/xxxx	xxxx

Appendix C – Display Table Sample for Aggregate Comparison of Death Sources, Section VII

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Algorithm 1 Total Without Ties	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
Exact	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
Class 1 total	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
Class 2 total	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
<34.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥34.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥39.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥44.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥49.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥54.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
Class 3 total	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
<27.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥27.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥32.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥37.5	999	999	999	999	999	999	999	999	999	99.90%	99.90%	99.90%	99.90%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Algorithm 2 Total Without Ties	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Exact	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 1 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 2 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥39.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥44.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥49.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥54.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
Class 3 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Algorithm 3 Total Without Ties	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Exact	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 1 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 2 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
<34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥34.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥39.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥44.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥49.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥54.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 3 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥47.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 4 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
<22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥22.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥27.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥32.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
≥37.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

	a (True Positive)	b (False Positive)	c (False Negative)	d (True Negative)	Test Positive (a+b)	Test Negative (c+d)	Death in Health Plan (a+c)	No Death in Health Plan (b+d)	Total population (a+b+c+d)	Sensitivity	Specificity	PPV	NPV
	N	N	N	N	N	N	N	N	N	%	%	%	%
≥42.5	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%
Class 5 total	999	999	999	999	999	999	999	999	999	99.9%	99.9%	99.9%	99.9%

Appendix D – Display Table Sample for Aggregate Incidence of Cardiovascular Death, Section VII

A2. Incidence of cardiovascular death (CVD) and all-cause mortality among users of antiarrhythmic medications and among non-users												
	Events per person and/or risk of CVD by patient characteristics		**Events per person and/or risk of all-cause mortality by patient characteristics**		**Number of CVD and death events, and denominators**							
Patient Factor	Incidence Antiarrhythmic med users (C1 Rate)	Incidence Non-users (Rate C2)	Incidence Antiarrhythmic med users (Rate C1)	Incidence Non-users (Rate C2)	C1 CVD events (Count)	C2 CVD events (Count)	C1 Death events (Count)	C2 Death events (Count)	C1 Denom All-cause mortality (Count)	C2 Denom All-cause mortality (Count)	C1 Denom CVD (Count)	C2 Denom CVD (Count)
Overall	CVD in C1/All patients in C1	CVD in C2/All patients in C2	Deaths in C1/All patients in C1	Deaths in C2/All patients in C2	CVD in C1	CVD in C2	Deaths C1	Deaths C2	All patients in C1	All patients in C2	All patients in C1 with NDI COD[7]	All patients in C2 with NDI COD[7]
Female	CVD in females in C1/All females in C1	CVD in females in C2/All females in C2	Deaths in females in AR / All females in C1	Deaths in females in C2/All females in C2	CVD in females C1	CVD in females C2	Deaths in females C1	Deaths in females C2	All female patients in C1	All female patients in C2	All female patients in C1 with NDI COD[7]	All female patients in C2 with NDI COD[7]
55-64 years	CVD in females 55-64 in C1/All females 55-64 in C1	CVD in females 55-64 in C2/All females 55-64 in C2	Deaths in females 55-64 in C1 / All females 55-64 in C1	Deaths in females 55-64 in C2 / All females 55-64 in C2	CVD in females 55-64 in C1	CVD in females 55-64 in C2	Deaths in females 55-64 in C1	Deaths in females 55-64 in C2	All females 55-64 in C1	All females 55-64 in C2	All female patients 55-64 years in C1 with NDI COD[7]	All female patients 55-64 years in C2 with NDI COD[7]

[7] Total is composed of only the following mutually exclusive groups:

1- C1 patient with no death

2- C1 patient with NDI death *and* NDI COD record

A2. Incidence of cardiovascular death (CVD) and all-cause mortality among users of antiarrhythmic medications and among non-users												
	Events per person and/or risk of CVD by patient characteristics		**Events per person and/or risk of all-cause mortality by patient characteristics**		**Number of CVD and death events, and denominators**							
Patient Factor	Incidence Antiarrhythmic med users (C1 Rate)	Incidence Non-users (Rate C2)	Incidence Antiarrhythmic med users (Rate C1)	Incidence Non-users (Rate C2)	C1 CVD events (Count)	C2 CVD events (Count)	C1 Death events (Count)	C2 Death events (Count)	C1 Denom All-cause mortality (Count)	C2 Denom All-cause mortality (Count)	C1 Denom CVD (Count)	C2 Denom CVD (Count)
65-74 years	CVD in females 65-74 in C1/All females 65-74 in C1	CVD in females 65-74 in C2 / All females 65-74 in C2	Deaths in females 65-74 in C1 / All females 65-74 in C1	Deaths in females 65-74 in C2 / All females 65-74 in C2	CVD in females 65-74 in C1	CVD in females 65-74 in C2	Deaths in females 65-74 in C1	Deaths in females 65-74 in C2	All females 65-74 in C1	All females 65-74 in C2	All female patients 65-74 years in C1 with NDI COD[7]	All female patients 65-74 years in C2 with NDI COD[7]
75-85 years	CVD in females 75-85 in C1 / All females 75-85 in C1	CVD in females 75-85 in C2 / All females 75-85 in C2	Deaths in females 75-85 in C1 / All females 75-85 in C1	Deaths in females 75-85 in C2 / All females 75-85 in C2	CVD in females 75-85 in C1	CVD in females 75-85 in C2	Deaths in females 75-85 in C1	Deaths in females 75-85 in C2	All females 75-85 in C1	All females 75-85 in C2	All female patients 75-85 years in C1 with NDI COD	All female patients 75-85 years in C2 with NDI COD
>85 years	CVD in females >85 in C1 / All females >85 in C1	CVD in females >85 in C2 / All females >85 in C2	Deaths in females >85 in C1 / All females >85 in C1	Deaths in females >85 in C2 / All females ages>85 in C2	CVD in females >85 in C1	CVD in females >85 in C2	Deaths in females >85 in C1	Deaths in females >85 in C2	All females >85 in C1	All females >85 in C2	All female patients >85 years in C1 with NDI COD	All female patients >85 years in C2 with NDI COD
Male	CVD in males in AR / All males in C1	CVD in males in C2/All males in C2	Deaths in males in AR / All males in C1	Deaths in males in C2/All males in C2	CVD in males C1	CVD in males C2	Deaths in males C1	Deaths in males C2	All male patients in C1	All male patients in C2	All male patients in C1 with NDI COD	All male patients in C2 with NDI COD

A2. Incidence of cardiovascular death (CVD) and all-cause mortality among users of antiarrhythmic medications and among non-users												
	Events per person and/or risk of CVD by patient characteristics		**Events per person and/or risk of all-cause mortality by patient characteristics**		**Number of CVD and death events, and denominators**							
Patient Factor	Incidence Antiarrhythmic med users (C1 Rate)	Incidence Non-users (Rate C2)	Incidence Antiarrhythmic med users (Rate C1)	Incidence Non-users (Rate C2)	C1 CVD events (Count)	C2 CVD events (Count)	C1 Death events (Count)	C2 Death events (Count)	C1 Denom All-cause mortality (Count)	C2 Denom All-cause mortality (Count)	C1 Denom CVD (Count)	C2 Denom CVD (Count)
45-54 years	CVD in males 45-54 in C1 / All males 45-54 in C1	CVD in males 45-54 in C2 / All males 45-54 in C2	Deaths in males 45-54 in C1 / All males 45-54 in C1	Deaths in males 45-54 in C2 / All males 45-54 in C2	CVD in males 45-54 in C1	CVD in males 45-54 in C2	Deaths in males 45-54 in C1	Deaths in males 45-54 in C2	All males 45-54 in C1	All males 45-54 in C2	All male patients 45-54 years in C1 with NDI COD	All male patients 45-54 in C2 with NDI COD
55-64 years	CVD in males 55-64 in C1 / All males 55-64 in C1	CVD in males 55-64 in C2 / All males 55-64 in C2	Deaths in males 55-64 in C1 / All males 55-64 in C1	Deaths in males 55-64 in C2 / All males 55-64 in C2	CVD in males 55-64 in C1	CVD in males 55-64 in C2	Deaths in males 55-64 in C1	Deaths in males 55-64 in C2	All males 55-64 in C1	All males 55-64 in C2	All female patients 55-64 years in C1 with NDI COD	All female patients 55-64 years in C2 with NDI COD
65-74 years	CVD in males 65-74 in C1 / All males 65-74 in C1	CVD in males 65-74 in C2 / All males 65-74 in C2	Deaths in males 65-74 in C1 / All males 65-74 in C1	Deaths in males 65-74 in C2 / All males 65-74 in C2	CVD in males 65-74 in C1	CVD in males 65-74 in C2	Deaths in males 65-74 in C1	Deaths in males 65-74 in C2	All males 65-74 in C1	All males 65-74 in C2	All female patients 65-74 years in C1 with NDI COD	All female patients 65-74 years in C2 with NDI COD
75-85 years	CVD in males 75-85 in C1 / All males 75-85 in C1	CVD in males 75-85 in C2 / All males 75-85 in C2	Deaths in males 75-85 in C1 / All males 75-85 in C1	Deaths in males 75-85 in C2 / All males 75-85 in C2	CVD in males 75-85 in C1	CVD in males 75-85 in C2	Deaths in males 75-85 in C1	Deaths in males 75-85 in C2	All males 75-85 in C1	All males 75-85 in C2	All female patients 75-85 years in C1 with NDI COD	All female patients 75-85 years in C2 with NDI COD

A2. Incidence of cardiovascular death (CVD) and all-cause mortality among users of antiarrhythmic medications and among non-users												
	Events per person and/or risk of CVD by patient characteristics		**Events per person and/or risk of all-cause mortality by patient characteristics**		**Number of CVD and death events, and denominators**							
Patient Factor	Incidence Antiarrhythmic med users (C1 Rate)	Incidence Non-users (Rate C2)	Incidence Antiarrhythmic med users (Rate C1)	Incidence Non-users (Rate C2)	C1 CVD events (Count)	C2 CVD events (Count)	C1 Death events (Count)	C2 Death events (Count)	C1 Denom All-cause mortality (Count)	C2 Denom All-cause mortality (Count)	C1 Denom CVD (Count)	C2 Denom CVD (Count)
>85 years	CVD in males >85 in C1 / All males >85 in C1	CVD in males >85 in C2 / All males >85 in C2	Deaths in males >85 in C1 / All males >85 in C1	Deaths in males >85 in C2 / All males ages>85 in C2	CVD in males >85 in C1	CVD in males >85 in C2	Deaths in males >85 in C1	Deaths in males >85 in C2	All males >85 in C1	All males >85 in C2	All female patients >85 years in C1 with NDI COD	All female patients >85 years in C2 with NDI COD
Comorbidities												
Coronary heart disease (CHD) (Denominator All patients with CHD present or absent)												
CHD presence	Patients with CHD and CVD, C1 / All patients with CHD C1	Patients with CHD and CVD, C2 / All patients with CHD C2	Patients with CHD and death, C1 / All patients with CHD C1	Patients with CHD and death, C2 / All patients with CHD C2	Patients with CHD and CVD, C1	Patients with CHD and CVD, C2	Patients with CHD and Death	Patients with CHD and Death	All Patients with CHD, C1	All Patients with CHD, C2	All Patients with CHD in C1, with NDI COD	All Patients with CHD in C2, with NDI COD
CHD absence	Patients with NO CHD and CVD, C1 / All patients with NO CHD C1	Patients with NO CHD and CVD, C2 / All patients with NO CHD C2	Patients with NO CHD and death, C1 / All patients with NO CHD C1	Patients with NO CHD and death, C2 / All patients with NO CHD C2	Patients with NO CHD and CVD, C1	Patients with NO CHD and CVD, C2	Patients with NO CHD and Death, C1	Patients with NO CHD and Death, C2	All Patients with NO CHD, C1	All Patients with NO CHD, C2	All Patients with NO CHD in C1, with NDI COD	All Patients with NO CHD in C2, with NDI COD
Diabetes Mellitus (DM) (Denominator All patients with DM present or absent)												

A2. Incidence of cardiovascular death (CVD) and all-cause mortality among users of antiarrhythmic medications and among non-users												
	Events per person and/or risk of CVD by patient characteristics		**Events per person and/or risk of all-cause mortality by patient characteristics**		**Number of CVD and death events, and denominators**							
Patient Factor	Incidence Antiarrhythmic med users (C1 Rate)	Incidence Non-users (Rate C2)	Incidence Antiarrhythmic med users (Rate C1)	Incidence Non-users (Rate C2)	C1 CVD events (Count)	C2 CVD events (Count)	C1 Death events (Count)	C2 Death events (Count)	C1 Denom All-cause mortality (Count)	C2 Denom All-cause mortality (Count)	C1 Denom CVD (Count)	C2 Denom CVD (Count)
DM presence	Patients with DM and CVD, C1 / All patients with DM C1	Patients with DM and CVD, C2 / All patients with DM C2	Patients with DM and death, C1 / All patients with DM C1	Patients with DM and death, C2 / All patients with DM C2	Patients with DM and CVD, C1	Patients with DM and CVD,C2	Patients with DM and Death, C1	Patients with DM and Death, C2	All Patients with DM, C1	All Patients with DM, C2	All Patients with DM in C1 with NDI COD	All Patients with DM in C2 with NDI COD
DM absence	Patients with NO DM and CVD, C1 / All patients with NO DM C1	Patients with NO DM and CVD, C2 / All patients with NO DM C2	Patients with NO DM and death, C1 / All patients with NO DM C1	Patients with NO DM and death, C2 / All patients with NO DM C2	Patients with NO DM and CVD, C1	Patients with NO DM and CVD, C2	Patients with NO DM and Death, C1	Patients with NO DM and Death, C2	All Patients with NO DM, C1	All Patients with NO DM, C2	All Patients with NO DM in C1, with NDI COD	All Patients with NO DM in C2, with NDI COD

Appendix E – Patient Characteristic Reference Table for Aggregate Incidence of Death and Death from Specific Causes, Section VII

Sort_Order	Group	Definition
01	Overall	All patients
02	Female	Sex=F
03	Female-55-64 years	Sex=F and Age between 55-64 inclusive
04	Female-65-74 years	Sex=F and Age between 65-74 inclusive
05	Female-75-85 years	Sex=F and Age between 75-85 inclusive
06	Female->85 years	Sex=F and Age greater than 85
07	Male	Sex=M
08	Male-45-54 years	Sex=M and Age between 45-54 inclusive
09	Male-55-64 years	Sex=M and Age between 55-64 inclusive
10	Male-65-74 years	Sex=M and Age between 65-74 inclusive
11	Male-75-85 years	Sex=M and Age between 75-85 inclusive
12	Male->85 years	Sex=M and Age greater than 85
13	CHD-Present	Patients with CHD, defined as Num_CHD=1
14	CHD-Absent	Patients without CHD, defined as Num_CHD=0
15	Diabetes-Present	Patients with diabetes, defined as Num_Diabetes=1
16	Diabetes-Absent	Patients without diabetes, defined as Num_Diabetes=0

Appendix F – Underlying cardiovascular cause of death diagnostic codes for Section VIII

Sample Codes With Decimals In Place

ICD_10	Description	Name1	Name2
I00	Rheumatic fever without mention of heart involvement	Major CVD	Rheumatic fever
I01.0	Acute rheumatic pericarditis	Major CVD	Rheumatic fever
I01.2	Acute rheumatic myocarditis	Major CVD	Rheumatic fever
I01.9	Acute rheumatic heart disease, unspecified	Major CVD	Rheumatic fever
I02.0	Rheumatic chorea with heart involvement	Major CVD	Rheumatic fever
I05.0	Mitral stenosis	Major CVD	Chronic rheumatic heart diseases
I05.2	Mitral stenosis with insufficiency	Major CVD	Chronic rheumatic heart diseases
I05.9	Mitral valve disease, unspecified	Major CVD	Chronic rheumatic heart diseases
I06.0	Rheumatic aortic stenosis	Major CVD	Chronic rheumatic heart diseases
I06.2	Rheumatic aortic stenosis with insufficiency	Major CVD	Chronic rheumatic heart diseases
I06.9	Rheumatic aortic valve disease, unspecified	Major CVD	Chronic rheumatic heart diseases
I07.0	Tricuspid stenosis	Major CVD	Chronic rheumatic heart diseases
I07.2	Tricuspid stenosis with insufficiency	Major CVD	Chronic rheumatic heart diseases
I07.9	Tricuspid valve disease, unspecified	Major CVD	Chronic rheumatic heart diseases
I08.0	Disorders of both mitral and aortic valves	Major CVD	Chronic rheumatic heart diseases
I08.2	Disorders of both aortic and tricuspid valves	Major CVD	Chronic rheumatic heart diseases

ICD_10	Description	Name1	Name2
I08.8	Other multiple valve diseases	Major CVD	Chronic rheumatic heart diseases

ATTACHMENT 3

Pilot study results presented at the 36th International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE) in 2020: A standardized and reusable method to link multiple distributed health plan databases to the National Death Index

A standardized and reusable method to link multiple distributed health plan databases to the National Death Index

Candace C. Fuller, MPH, PhD[1]; Wei Hua, MD, PhD, MSc, MHS[2]; Robert Ball, MD, MPH, ScM[2]; Ryan Carnahan, PharmD, MS[3]; Michael Dizinno, MPH, M.ED[1]; Sarah Dutcher, PhD, MS[2]; Charles E. Leonard, PharmD, MSCE[4]; Andrew Mosholder, MD, MPH[2]; Andrew B. Petrone, MPH[1]; Robert Rosofsky, MA[5]; Laura Shockro[1]; Jessica Young, PhD[1]; Margaret Adgent, MSPH, PhD, PharmD, MPH[8]; Ingrid Binswanger, MD, MPH, MS[6]; Denise Boudreau, RPh, PhD, MS[7]; Marie Griffin, MD, MPH[8]; Jennifer Kuntz, MS, PhD[9]; Cheryl McMahill-Walraven, MSW, PhD[10]; Pamala A. Pawloski, PharmD[11]; Sengwee Toh, ScD[1]

[1]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [2]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD; [3]University of Iowa, College of Public Health Iowa City, IA; [4]Center for Pharmacoepidemiology Research and Training, Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; [5]Health Information Systems Consulting, Milton, MA; [6]Institute for Health Research Kaiser Permanente Colorado, Aurora, CO; [7]Kaiser Permanente Washington Health Research Institute Seattle, WA; [8]Vanderbilt University Medical Center, Nashville, TN; [9]Kaiser Permanente Northwest, Portland, OR; [10]Aetna, Blue Bell, PA; [11]HealthPartners Institute, Bloomington, MN

Disclosures

No conflicts of interest. This work was supported by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223301710132C. This abstract reflects the views of the authors and should not be construed to represent FDA's views or policies.

Background

- In the US, health plan (HP) databases typically capture medically attended deaths during enrollment, but many do not capture out-of-hospital deaths.
- Linkage to the National Death Index (NDI) can augment HP-based studies with death and cause of death information, but linkage for multi-site studies can be resource intensive.

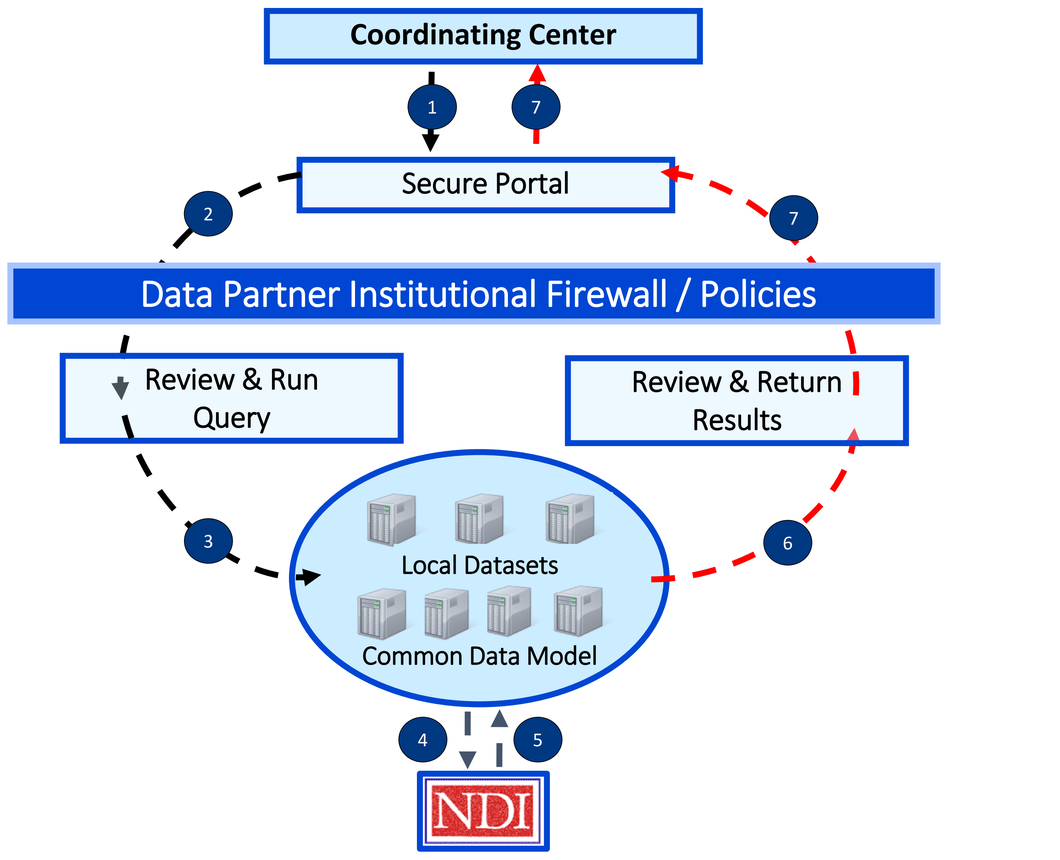
Objective

- To develop and test a reusable distributed NDI linkage approach within a use case comprising antiarrhythmic (AAD) drug users and non-users.

Methods

- As a test case, we identified a cohort of AAD users, known to be at increased risk of death, in 6 HPs in years 2000-2017 and created exposure episodes using dispensing data.
- For reference, we matched AAD users 1:1 on age, sex, and HP with non-users.
- We followed AAD users and matched non-users for death from the AAD user's index dispensing date.
- We submitted patients with any of the following to the NDI for linkage: (a) HP-recorded death during the exposure episode plus 365 days, (b) potential death, or (c) incomplete follow-up (i.e., HP data ended before patients could be followed for the entire exposure episode plus 365 days).
- We defined potential death as HP disenrollment during the exposure episode plus 365 days, without reenrollment or medical utilization >60 days post disenrollment from the HP.

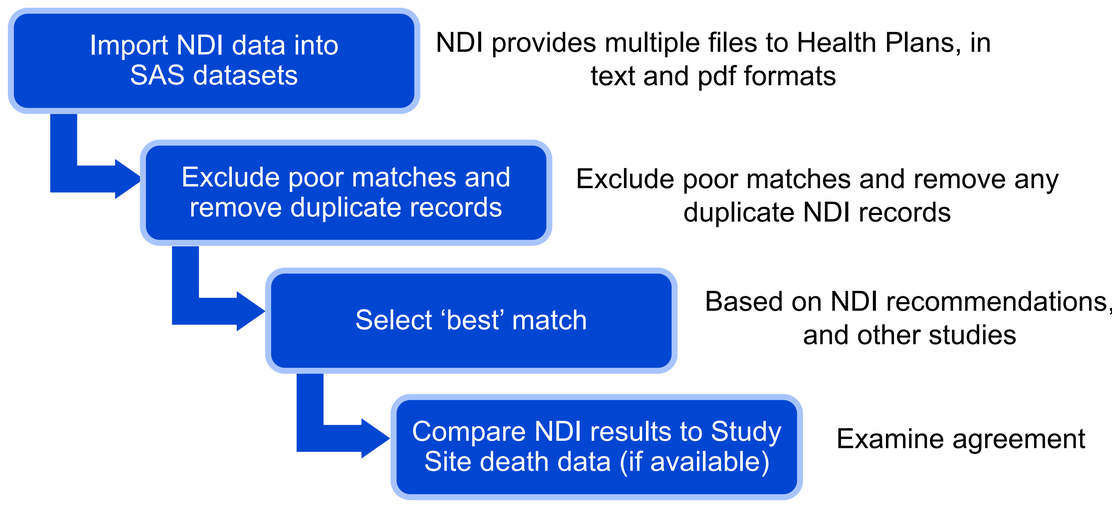
Figure 1. Overview of the Distributed National Death Index (NDI) Linkage Process



1 The study team initiates a query identifying patients of interest
2 Participating health plans retrieve the query on the secure portal
3 Participating health plans review and run query for identifying patients to submit to the NDI
4 Participating health plans prepare files to submit to the NDI
5 The NDI returns files to participating health plans
6 Participating health plans run programs developed by the study team against returned NDI files to identify matches to be saved
7 Participating health plans remove all protected health information (PHI), save other data to protect analytical files, and return results for review by the study team

- We developed and tested reusable distributed NDI linkage programs which avoid sharing identifiable patient information between HPs or with the coordinating center (Figure 1).
- Distributed programs identified cohorts, selected potential deaths for NDI submission, quality checked NDI records flagged for submission, and selected the single best match from returned NDI data (Figure 2).

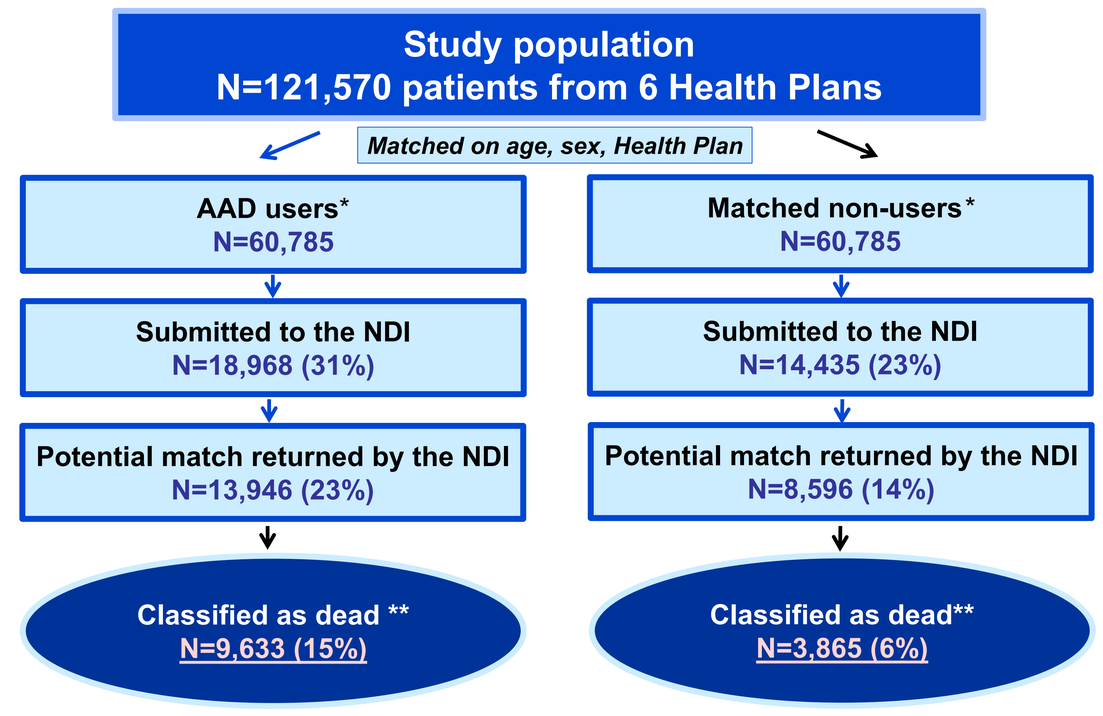
Figure 2. General process for receiving NDI data and selecting the best match



- At 5 HPs that ascertain death with state death records and other sources we examined concordance, sensitivity, and positive predictive value of death identified through NDI linkage compared to HP-recorded death (reference).

Results

Figure 3. Patients submitted to the National Death Index, returned with a possible match, and ultimately classified as dead by study algorithm



*Antiarrhythmic drug users
**Classification was completed by a program for selecting the single best NDI match from all potential matches provided by the NDI

- The NDI returned at least one potential match for 22,542 (67%) patients submitted to the NDI.
- Our program for selecting the best NDI match classified 13,498 (40%) patients submitted to the NDI as dead.

Table 1. Concordance between Health Plan (HP) recorded deaths and deaths identified with data from the National Death Index (NDI), at 5 HPs

Total patients, N=12,956			
	HP Recorded Death=Yes	HP Recorded Death=No	Total
NDI Death=Yes	7,402	238	7,640
NDI Death=No	269	5,047	5,316
Total	7,671	5,285	

- We observed excellent concordance between HP-recorded deaths and NDI-identified death (96% agreement; Cohen's Kappa=0.919).
- The NDI captured 7,402 of 7,671 HP-recorded deaths (Sn=97%, PPV=97%).
- The NDI identified 238 deaths not identified in HP data; review of discordant records is ongoing.

Conclusions

- The developed approach standardized the NDI linkage across HPs.
- We observed high concordance between the best match chosen by the NDI linkage program and HP-recorded death.
- Future work will further test and refine the reusable distributed NDI linkage approach, with the goal of providing NDI linkage methods to future studies based in distributed data networks.