Redacted Protocol (Pangea ID 2024-14140) Development of Algorithms to Identify Intracerebral Hemorrhage Greater Than 1 cm and Amyloid-Related Imaging Abnormalities Using Electronic Medical Records in Select Populations of Alzheimer's Disease and Related Dementias and Mild Cognitive Impairment Patients in the United States

Protocol Electronically Signed and Approved by Lilly: approval date provided below

Approved on 27 Jun 2025 GMT

Marketing Authorisation Holder

Marketing Authorisation Holder (MAH)	Eli Lilly Global Patient Safety Pharmacoepidemiologist Eli Lilly and Company Lilly Corporate Center Indianapolis, IN 46285
MAH Contact Person	Lilly Global Patient Safety Pharmacoepidemiologist

1. Table of Contents

Section **Page**

List of Tables

List of Figures

List of Annexes

Annex **Page**

2. List of Abbreviations

Term	Definition
AChEI	acetylcholinesterase inhibitor
AD	Alzheimer's disease
ADRD	Alzheimer's disease and related dementias
AE	adverse event
ARIA	amyloid-related imaging abnormalities
ARIA-E	amyloid-related imaging abnormalities - edema or effusion
ARIA-H	amyloid-related imaging abnormalities - microhemorrhage(s) or superficial siderosis
ATT	amyloid-targeting therapies
eCRF	electronic case report form
CMS	Centers For Medicare and Medicaid Services
CT	computed tomography
EMR	electronic medical record
FFS	fee-for-service
HIE	health information exchange
HIPAA	Health Insurance Portability and Accountability Act
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ICH	intracerebral hemorrhage
ID	identifier
MCI	mild cognitive impairment
MRI	magnetic resonance imaging
NLP	natural language processing

Term	Definition
NPV	negative predictive value
PPV	positive predictive value
SME	subject matter expert
TBI	traumatic brain injury
VDW	visit-data warehouse

3. Responsible Parties

Sponsor:

Eli Lilly and Company,

Lilly Corporate Center,

Indianapolis,

IN 46285, USA

Third Party Vendor:

Premier Healthcare Solutions, Inc.

113034 Ballantyne Corporate Place

Charlotte, NC 28277

Contact details and list of investigators are in a standalone document available upon request.

4. Abstract

- Title: Development of algorithms to identify intracerebral hemorrhage greater than 1 cm (ICH >1 cm) and amyloid-related imaging abnormalities (ARIA) using electronic medical records in select populations of Alzheimer's disease and related dementias and patients with mild cognitive impairment in the US.
- Rationale and background: Patients receiving amyloid-targeting therapies (ATTs) need to be monitored for ICH >1 cm and ARIA. There are currently no validated algorithms to assess these outcomes in secondary databases. Developing algorithms for use with electronic medical records of head computed tomography (CT) and brain magnetic resonance imaging (MRI) scans would facilitate identification and monitoring of events.
- Objectives: 1) Develop a natural language processing (NLP) algorithm using head CT and/or brain MRI reports of patients with Alzheimer's disease and related dementias or mild cognitive impairment to identify ICH >1 cm and estimate the algorithm's performance with manual review of imaging reports. 2) Develop NLP algorithms using brain MRI reports of patients treated with ATTs to identify ARIA-E (ARIA with edema or effusion) and ARIA-H (ARIA with superficial siderosis or microhemorrhage(s)) and estimate the algorithms' performances with manual review of imaging reports. 3) Describe patient demographics, medication use, and comorbidities for each group of reports used for algorithm development and validation. 4) Identify and characterize potential comparator cohorts for future safety assessment of ATTs.
- Study design: Observational study using secondary data to develop and validate NLP algorithms. The data will include structured codes and terms (such as diagnoses, procedures, and medications) and unstructured narratives from CT and MRI reports from electronic medical records.
- Population: Reports from patients aged 65 years and older with an inpatient or outpatient visit between 01 January 2022 through the latest data available and a head CT and/or brain MRI report. Reports limited to patients treated with ATTs for ARIA algorithms development. Patients for comparator cohorts will be determined based on Alzheimer's disease and related dementias or mild cognitive impairment diagnosis, ATTs, and acetylcholinesterase inhibitors treatment.
- Variables: ICH >1 cm, ARIA-E, ARIA-H (including delineation by subtype), type of imaging report, and patient characteristics (i.e., age, sex, race, ethnicity, primary insurance, comorbid conditions, use of antithrombotic medication, and traumatic brain injury diagnosis). Use of ATTs and acetylcholinesterase inhibitor treatment will be evaluated as applicable for inclusion/exclusion criteria.
- Data sources: PPD Health Information Exchange (HIE) Data.

- Study size: All imaging reports meeting applicable inclusion criteria will be included in the development of the NLP algorithms. Imaging reports from 400 patients will be included in the validation step. Comparator cohorts will be comprised of all patients meeting specified criteria.
- Data analysis: NLP algorithms for ICH >1 cm, ARIA-E, and ARIA-H will be developed and refined. Performance measures will be used to summarize the results of NLP algorithms validation (i.e., sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and F1 scores). Descriptive analysis will be used to characterize patients whose reports were used for algorithm development and validation, as well as the potential comparator cohorts.
- Milestones: Secondary data of existing records will be obtained from 01 January 2022 through the latest data available. Completion of the final study report is planned for 30 April 2027.

5. Amendments and Updates

Not applicable.

6. Rationale and Background

Alzheimer's disease and related dementias (ADRD) and mild cognitive impairment (MCI) are among the most common chronic conditions in older adults. It has been estimated that 6.9 million individuals over age 65 in the United States (US) are living with Alzheimer's disease (AD) (Alzheimer's Association 2024; Rajan et al. 2021). The prevalence of MCI in people aged 65 and older has been estimated around 17% to 22% (Alzheimer's Association 2024; Manly et al. 2022; Petersen et al. 2018). As global life expectancy increases, the percentage of people with ADRD and MCI is expected to increase (Alzheimer's Association 2024; Rajan et al. 2021). Specifically, it is estimated that the number of people with AD and MCI will continue to increase by 18% and 9%, respectively, over the next 40 years (Rajan et al. 2021). AD is also associated with increased mortality and is currently the sixth-leading cause of death in the US (Alzheimer's Association 2024; US-DHSS 2023).

AD is characterized by the buildup of beta-amyloid and tau proteins, which accumulate outside and inside neurons, respectively. This protein accumulation causes neurodegeneration or damage and destruction of brain cells (Hampel et al. 2021; Ma et al. 2022). Studies assessing biomarkers for AD with positron emission tomography scans report that about half of people with MCI have AD-related brain changes (Alzheimer's Association 2024; Petersen et al. 2013; Rabinovici et al. 2019).

Amyloid-targeting therapies (ATTs), such as FDA-approved anti-amyloid monoclonal antibodies (MABs) donanemab (Kisunla™) and lecanemab (Leqembi™), target amyloid plaques in the brain and slow the progression of dementia (Sims et al. 2023; van Dyck et al. 2023). During treatment with these ATTs, patients need to be monitored for occurrence of AEs such as ARIA, ICH, and infusion reactions (Kisunla prescribing information, 2024). ARIA is an imaging abnormality that occurs predominantly in people treated with ATTs, including donanemab. Nevertheless, cases of ARIA-H with microhemorrhages have been shown to occur spontaneously in up to 32% of patients with AD, whereas spontaneous cases of ARIA-E may rarely occur (Hampel et al. 2023; Zimmer et al. 2025). ARIA is often asymptomatic with the only indication of the AEs observed via MRI. ARIA is commonly observed as ARIA-H (microhemorrhage(s) or superficial siderosis characterized by hemosiderin deposits) or ARIA-E (temporary swelling in an area or areas of the brain with vasogenic edema or sulcal effusion). Both ARIA-E and ARIA-H are detected by MRI. The ARIA events may be serious and even fatal in some cases.

In donanemab placebo-controlled clinical trials, ARIA-E (asymptomatic or symptomatic) was reported in 24.4% of 984 patients treated with donanemab. Symptomatic ARIA-E was reported in 5.8% of donanemab-treated patients, and serious ARIA-E was reported in 1.5%. In placebo-controlled clinical trials, ARIA-H (asymptomatic or symptomatic) was reported in 31.3% of patients treated with donanemab, symptomatic ARIA-H was reported in 1.0% and serious ARIA-H was reported in 0.4% of patients treated with donanemab. Separately, ICH >1 cm was reported in 3 patients treated with donanemab (0.3%) and serious ICH >1 cm was reported in 1

patient (0.1%) of patients treated with donanemab (Eli Lilly and Company [Lilly] internal data; (Kisunla prescribing information, 2024) (Zimmer et al. 2025)). Given the recommendations for monitoring of ARIA-E and ARIA-H in treated patients (i.e., obtain a recent baseline brain MRI prior to initiating treatment; obtain an MRI prior to the second, third, fourth, and seventh infusion, see Kisunla prescribing information, 2024), ARIA events are expected to be detected in a larger proportion among patients treated with ATTs.

Higher risk of ARIA was demonstrated in clinical trial patients with the apolipoprotein E ε4/ε4 genotype, which is also a known risk factor for the development of AD and cerebral amyloid angiopathy. Additional risks for ARIA include the presence of baseline microhemorrhage(s) or cortical superficial siderosis, higher brain amyloid burden, as well as higher systemic blood pressure (Zimmer et al. 2025). The presence of these risk factors in donanemab users overall and those experiencing ARIA events will help develop hypotheses about the impact of these risk factors in real world populations.

ICH >1 cm with donanemab treatment in placebo-controlled clinical trials was infrequent (0.3%). However, serious (including fatal) cases of ICH >1 cm with concomitant use of thrombolytics or anticoagulant medications have been observed with ATT, warranting the continued assessment of this potential AE in the postmarketing setting (internal Lilly advisory committee briefing document).

Patients with ADRD have an increased risk of ICH independent of ATT treatment. In patients with AD, a systematic review and meta-analysis of 20 studies totaling 61,824 patients showed that the incidence rates were 15.4/1000 person-years for stroke (all types), 13.0/1000 person-years for ischemic stroke, and 3.4/1000 person-years for ICH (Pinho et al. 2021). Compared to controls without AD, the incidence rate for ICH in patients with AD was significantly higher (incident rate ratio = 1.67, 95% confidence interval: 1.43-1.96), but similar for ischemic stroke. Older studies indicate that cognitive impairment prior to ICH in patients with primary ICH in any location is common (Cordonnier et al. 2010; Rost et al. 2008; Xiong et al. 2016). One study reported a prior cognitive impairment incidence of 15% in 629 patients with ICH (Rost et al. 2008), while another showed 14% of patients had preexisting cognitive impairment without dementia, and 16% had preexisting dementia (Cordonnier et al. 2010).

Understanding hematoma size is integral to evaluating severity and prognosis after ICH (LoPresti et al. 2014). CT or MRI is considered first-choice imaging option for assessment of ICH (Kidwell and Wintermark 2008). In the current study using natural language processing (NLP algorithms, the cutoff measure of >1 cm (or 10 mm) in any direction has been agreed with regulators (i.e., FDA and Medicines and Healthcare Products Regulatory Agency) as a proxy for notable ICH size.

Currently, the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)has no specific diagnosis codes for ICH >1 cm, ARIA-E, or ARIA-H. Therefore, a review of imaging reports is necessary to provide clinical details and retrieve relevant

information. ICH is detected in head CT scans or brain MRI scans, and ARIA events are detected in brain MRI scans (most often in patients treated with ATTs). Development and validation of algorithms to identify patients with ICH >1 cm, ARIA-E, and ARIA-H in a real-world data source will facilitate monitoring of these important ATT risks in large real-world populations.

To this end, this study will attempt to develop and validate NLP algorithms to detect the following events of interest in EMRs at scale: ICH >1 cm, ARIA-E, and ARIA-H. NLP algorithms will be developed and validated for future application with large real-world EMR data sets, allowing for the identification of events faster and in larger volume than can be achieved with systematic manual review of imaging reports.

To better understand the reports on which the NLP algorithms are developed and validated, general demographics, medication use, and specific comorbid conditions will be assessed in patients who contribute reports. Risk factors associated with developing ICH or ARIA outside of receiving ATT are of particular interest. Similarly, comorbidities influencing patients' propensity to get brain imaging or potentially influencing the interpretation of MRI /CT will be assessed. Therefore, traumatic brain injury (TBI, which has an estimated incidence of 13% in people aged more than 65 years old (Kornblith et al. 2024) and could result in ICH, will be appraised via sensitivity analysis. Patient groups with other brain pathology diagnoses (e.g., ischemic stroke) will also be considered for sensitivity analysis. Medications assessment will include antidementia and anti-amyloid treatment, as well as antithrombotic medications, which are characteristic in patients with ADRD and MCI.

Ultimately, the algorithms developed in this study will support Lilly's goal of assessing the frequency of ICH >1 cm and ARIA events in patients with ADRD or MCI undergoing donanemab treatment, using real-world data from CT and MRI scan reports in EMRs. Additionally, as part of a postmarketing requirement by FDA, this study will identify and descriptively characterize potential comparator cohorts of patients to support a potential forthcoming comparative safety study between donanemab-treated patients and comparator cohorts using secondary data. Similarly, to support a potential forthcoming study, published EMR based-algorithms or methods to identify other safety outcomes of interest (seizure, anaphylaxis, and death) will be described.

Lilly contracted with Premier Healthcare Solutions to execute the study, develop NLP algorithms, perform medical chart review for algorithm validation, and complete preparation activities for a potential future study.

7. Research Questions and Objectives

7.1. Research Questions

The study will address the following research questions:

1. What is the performance of an NLP algorithm, developed from head CT and/or brain MRI imaging reports of patients with ADRD or MCI, in identifying ICH greater than one centimeter (>1 cm or >10 mm)?

2. What are the performances of NLP algorithms, developed from brain MRI imaging reports of patients treated with ATTs, in identifying ARIA-E, and ARIA-H subtypes of ARIA-H with superficial siderosis and ARIA-H with microhemorrhage(s)?

7.2. Study Objectives

The purpose of the study is to set up a framework to use EMRs to identify imaging reports with the events of interest, as well as characterize comparator cohorts to allow meaningful safety assessment of ATTs in the future.

To meet the following study objectives, different groups of imaging reports will be defined and utilized for algorithms development and validation (Section 8.2.1).

The primary objective of the study is to

- Develop an NLP algorithm using CT and/or MRI reports of patients with ADRD or MCI to identify ICH >1 cm and estimate the algorithm's performance with manual review of imaging reports.

The secondary objectives include the following:

- Develop NLP algorithms using MRI reports of patients treated with ATTs to identify ARIA-E and ARIA-H subtypes (i.e., ARIA-H with superficial siderosis, ARIA-H with microhemorrhage(s)) and estimate the algorithms' performances with manual review of imaging reports.
- Based on data from structured EMRs, describe patient demographics, medication use, and comorbidities for each group of reports used for ICH >1 cm or ARIA algorithm development and validation.
- Identify and characterize potential comparator cohorts of patients for future safety assessment of ATTs.
- Describe published EMR-based algorithms or alternative methods that could be used to identify seizure, anaphylaxis, and death in a future safety assessment of ATTs.

7.3. Hypothesis

There is no statistical hypothesis for this study as the objectives are to develop and validate NLP algorithms, describe patient characteristics of the imaging report groups, identify and characterize potential comparator cohorts of patients, and describe published EMR-based methods for other safety outcomes. No hypothesis testing will be performed, and only descriptive statistics will be provided.

8. Research Methods

8.1. Study Design

This observational study will develop and validate NLP algorithms for ICH >1 cm, ARIA-E, and ARIA-H using EMRs data. The EMR data will include structured codes and terms (such as diagnoses, procedures, and medications) and unstructured CT and MRI report data from EMRs across multiple health systems, extracted from a Health Information Exchange (HIE) database, and de-identified in accordance with Health Insurance Portability and Accountability Act (HIPAA) regulations.

As part of development and validation of the NLP algorithms, the study aims to identify ICH >1 cm, ARIA-E, and ARIA-H in imaging report narratives, using NLP for algorithm development and manual review for validation. Measures of algorithm performances will be calculated for validation and will include sensitivity, specificity, PPV, NPV, accuracy, and F1 scores. Descriptive measures (e.g., frequency) of patient demographic and clinical characteristics (e.g., age, race, comorbid conditions, and antithrombotic medication use) will be summarized separately for each group of reports used for algorithm development and validation.

In addition to algorithm development and validation, this study also includes the following 2 activities that will be implemented in preparation for a potential future study of donanemab safety using EMR data:

1 identify and descriptively characterize potential comparator cohorts for future safety assessment of ATTs using EMR data, and

2 literature review aimed at documenting EMR-based algorithms or methods used to identify other safety outcomes of interest (seizure, anaphylaxis, and death).

8.2. Setting

The study will leverage data from PPD the largest HIE database in the US, which captures a diverse sample of providers and patients. It will include patients treated in inpatient and outpatient settings, as the data source includes hospital and clinic EMR data from over 9000 healthcare facilities, allowing for a representative sample of patients with ADRD and MCI.

8.2.1. Study Population

8.2.1.1. Identification of Reports for the Development of an NLP Algorithm to Identify ICH >1 cm (Report Group 1)

The development of the NLP algorithm for ICH >1 cm will occur at the report level. All imaging reports (CTs or MRIs) meeting the report inclusion/exclusion criteria will be included, and the entry event will be the report date (head CT or brain MRI) such that there is 1 row for each

report. Therefore, a single patient may have multiple study-eligible reports during the study period.

The study population used to develop an NLP algorithm to identify ICH >1 cm (i.e., Report group 1) will meet the following inclusion and exclusion criteria:

- **Patient inclusion criteria (for Report Group 1):**
  - Patients in the PPD HIE database during the study period from 01 January 2022 through the latest data available
  - With at least 1 head CT and/or brain MRI report available in the study period
- **Patient exclusion criteria (for Report Group 1):**
  - None
- **Report Group 1 inclusion criteria:**
  - Met patient inclusion/exclusion criteria above
    - Entry event = day of imaging report (patients contribute all head CT and/or brain MRI reports during the study period)
  - Patient aged 65 years or older at the entry event
  - Patient diagnosed with ADRD and/or MCI (diagnoses listed in Standalone Document No. 1) any time before or on the day of the entry event
- **Report Group 1 exclusion criteria:**
  - None

8.2.1.2. Identification of Reports for the Development of NLP Algorithms to Identify ARIA (Report Group 2)

The development of the NLP algorithms for ARIA events will occur at the report level. All imaging reports (brain MRIs) meeting the report inclusion/exclusion criteria will be included and the entry event will be the report date (brain MRI) such that there is 1 row for each report. Therefore, a single patient may have multiple imaging reports eligible for assessment during the study period.

Since ARIA is a specific safety outcome associated with ATT treatment, and because of the recommendations for monitoring of ARIA-E and ARIA-H in treated-patients (i.e., obtain a recent baseline brain MRI prior to initiating treatment; obtain an MRI prior to the second, third, fourth, and seventh infusion, see Kisunla prescribing information, 2024), routine MRIs are expected to influence the diagnosis of ARIA. Therefore, as patients without ATT use are not expected to receive an ARIA diagnosis, brain MRIs from patients treated with ATT will be utilized for ARIA algorithm development.

The study population used to develop NLP algorithms to identify ARIA (i.e., Report Group 2), including ARIA-E, and ARIA-H subtypes of ARIA-H with superficial siderosis, and ARIA-H with microhemorrhage(s), will meet the following inclusion and exclusion criteria:

- **Patient inclusion criteria (for Report Group 2):**
  - Patients in the PPD HIE database during the study period from 01 January 2022 through the latest data available
  - With at least 1 brain MRI report available in the study period
  - Patient with evidence of receiving an ATT (i.e., donanemab, lecanemab, and aducanumab as listed in Table 1) any time during the study period
- **Patient exclusion criteria (for Report Group 2):**
  - None
- **Report Group 2 inclusion criteria:**
  - Met patient inclusion or exclusion criteria above
    - Entry event = day of imaging report (patients contribute all MRI reports during the study period)
  - Patient aged 65 years or older at the entry event
- **Report Group 2 exclusion criteria:**
  - None

Both baseline (i.e., prior to receiving the first dose of ATT medication) and follow-up MRI reports will be included but will include no information on temporality between the reports and ATT treatment.

Table 1. Amyloid-Targeting Therapies

Nonproprietary Name	Proprietary Name	Approval Date	Withdrawal Date
Donanemab	Kisunla	02 Jul 2024	Not Applicable
Lecanemab	Leqembi	06 Jan 2023	Not Applicable
Aducanumab	Aduhelm	07 Jun 2021	01 Nov 2024

8.2.1.3. Reports Utilized for NLP Algorithms Validation

To facilitate validation of the NLP algorithms, manual chart review of imaging reports from subsets of patients utilized for algorithm development will be executed. A total of 400 patients will be selected for inclusion in the validation subgroups: 198 patients from Report Group 1 utilized for development of the NLP algorithm to identify ICH >1 cm, and 202 patients from Report Group 2 utilized for the development of NLP algorithms to identify ARIA (Sample Size Section 8.5).

Of important note, a separate study is attempting to develop a claims-based algorithm for ICH >1 cm and ARIA events, as part of a Kisunla FDA postmarketing requirement. In an effort to enable discussions regarding the performances of ICH >1 cm and ARIA algorithms across this study and the claims-based algorithms study, Lilly will utilize the same validation subgroups for both studies. This requires the PPD data to be linked to Medicare claims. To facilitate this, 2

patient selection criteria will be implemented prior to random selection of 400 patients for the validation step. These criteria aim to increase the likelihood that the patients selected into the validation subgroups can be linked to the Medicare claims data.

These criteria are as follows:

- First, given the lag in Medicare Advantage data, the claims-based algorithms will be developed on Medicare FFS patients only. Therefore, the validation subgroups will be limited to patients with FFS insurance, as identified in PPD EMR data.
- Second, linkage between the charts and Medicare claims will be facilitated through personal identifiable information available from a third party, PPD . Hence, patients selected for the validation subgroups will additionally be limited to those who have a linked PPD Study ID to PPD ID.

Once these 2 criteria are applied, selection of the final 400 patients for the validation subgroups will be executed as described in the Report Group 1 Validation Subgroup and Report Group 2 Validation Subgroup descriptions below. The process flow for the identification of PPD matched PPD patients is described in Section 8.6.2, Data Management.

All imaging reports meeting the inclusion/exclusion criteria from these randomly selected patients will be included in the validation phase. As patients on ATT are expected to receive more routine MRIs, this subgroup split ensures adequate report number for the validation of the algorithm to identify ICH >1 cm and algorithms to identify ARIA.

Report Group 1 Validation Subgroup will be selected as follows:

- **Patient inclusion criteria (for Report Group 1 Validation Subgroup):**
  - Met patient inclusion/exclusion criteria for "Report Group 1"
  - With FFS coverage (as identified in EMR data) anytime during the study period (head CT or brain MRI)
  - Patient present in PPD datafile
  - Evidence of a nontraumatic ICH determined via ICD-10-CM diagnosis in the structured data (see Standalone Document No. 3) during the study period*
- **Patient exclusion criteria (for Report Group 1 Validation Subgroup):**
  - None

NOTE: Patients meeting the inclusion criteria above will be randomly selected for the following inclusion/exclusion criteria.

- **Report Group 1 Validation Subgroup inclusion criteria:**
  - Met patient inclusion/exclusion criteria above
    - Entry event = day of imaging report (patients contribute all CT and/or MRI reports during the study period)
  - Patient aged 65 years or older at the entry event

- **<u>Report Group 1 Validation Subgroup exclusion criteria</u>:**
  - None

*Patients in the Report Group 1 Validation Subgroup are required to have evidence of a nontraumatic ICH event to increase the proportion of positive cases in this validation subgroup.

For the 198 patients in the Report Group 1 Validation Subgroup, all eligible head CT and brain MRI reports will be included. If most patients with ICH diagnosis present an ICH >1 cm, it is expected that the NLP algorithm will identify at least 1 imaging report with the outcome ICH >1 cm for each patient in Report Group 1 validation subgroup. Data from the Report Group 1 Validation Subgroup will be used to evaluate the NLP algorithm's performances in ICH >1 cm (in combination with Report Group 2 Validation Subgroup).

Report Group 2 Validation Subgroup will be selected as follows:

- **<u>Patient inclusion criteria (for Report Group 2 Validation Subgroup)</u>:**
  - Met patient inclusion/exclusion criteria for "Report Group 2"
  - With FFS coverage (as identified in PPD EMR data) anytime during the study period
  - Patient present in PPD datafile
- **<u>Patient Exclusion Criteria (for Report group 2 Validation Subgroup)</u>:**
  - None

NOTE: 202 patients meeting the inclusion criteria above will be randomly selected for the following inclusion/exclusion criteria.

- **<u>Report Group 2 Validation Subgroup inclusion criteria:</u>**
  - Met patient inclusion/exclusion criteria above
    - Entry event = day of imaging report (patients contribute all MRI reports during the study period)
  - Patient aged 65 years or older at the entry event
- **<u>Report Group 2 Validation Subgroup exclusion criteria:</u>**
  - None

All eligible MRI reports during the study period will be included, but no information on temporality between the reports and treatment will be available. Assuming that 24% of patients receiving ATT experience ARIA-E (Sims et al. 2023), we expect to have at least 1 brain MRI imaging report identified with ARIA-E by the developed NLP algorithm for about 24 patients in Report Group 2 Validation Subgroup. Assuming that 31% of patients receiving ATT experience ARIA-H (Sims et al. 2023), we expect to have at least 1 brain MRI imaging report identified with ARIA-H by the developed NLP algorithm for about 31 patients in Report Group 2 Validation Subgroup. Data from the Report Group 2 Validation Subgroup will be used to evaluate the NLP algorithms performances in identifying ARIA-E, and ARIA-H subtypes of

ARIA-H with superficial siderosis, and ARIA-H with microhemorrhage(s), as well as ICH >1 cm (in combination with Report Group 1 Validation Subgroup).

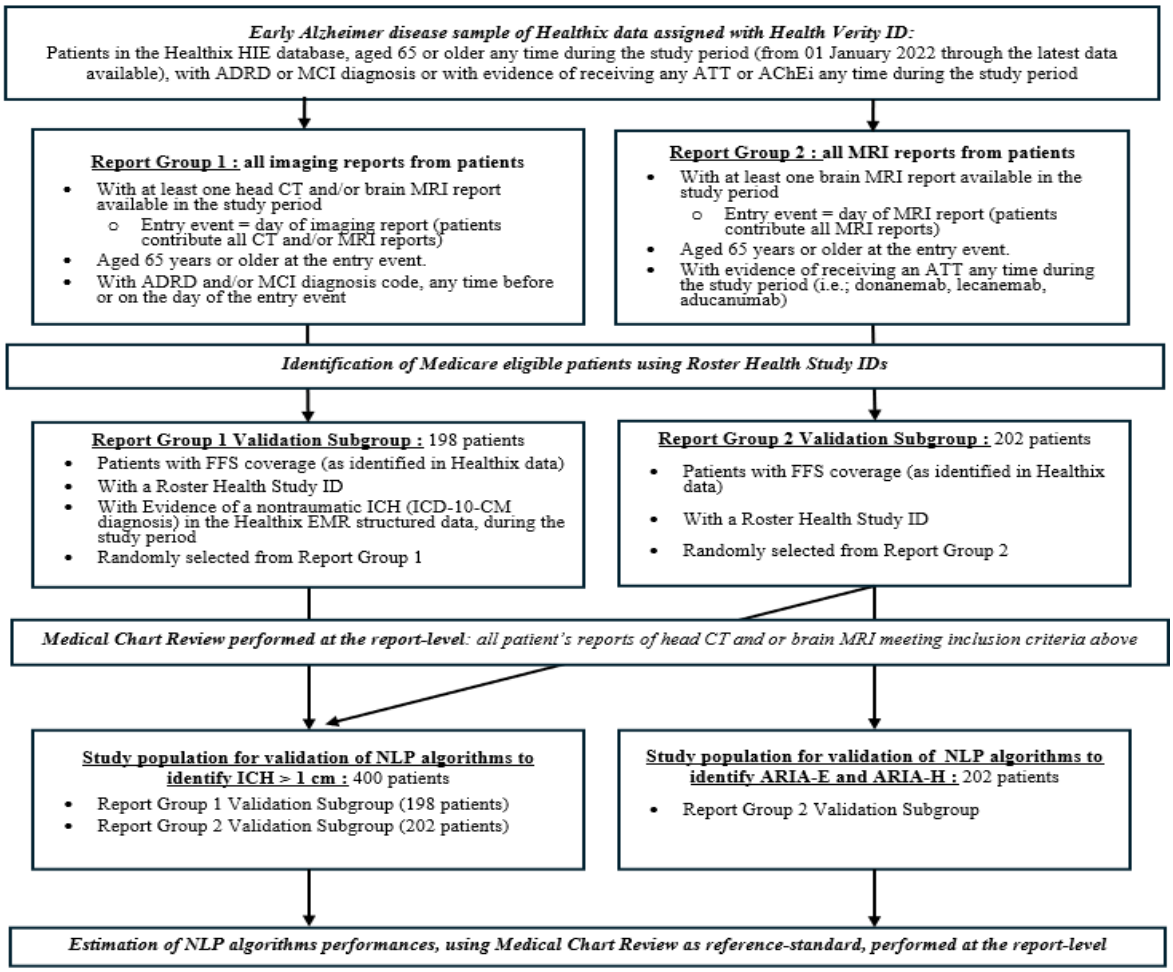
Validation of the NLP algorithm developed to identify ICH >1 cm:

The validation of the NLP algorithm to identify ICH >1 cm will be performed using all eligible reports (head CT and brain MRIs) from a maximum pooled group of 400 patients, composed of Report Group 1 Validation Subgroup (198 patients) in combination with Report Group 2 Validation Subgroup (202 patients). The Report Group 2 Validation Subgroup is being added to ensure an adequate number of negative ICH imaging reports. A single patient may contribute multiple reports during the study period. The validation of the NLP algorithm will occur at the report level.

Validation of the NLP algorithms developed to identify ARIA:

The validation of the NLP algorithms to identify ARIA will be performed using data from the Report Group 2 Validation Subgroup only (202 patients). A single patient may contribute multiple reports during the study period. The validation of the NLP algorithms will occur at the report level.

The flow diagram of the validation study, including the selection of imaging report for medical review, is depicted in Figure 1.



Abbreviations: AChEI = acetylcholinesterase inhibitor; ADRD = Alzheimer's disease and related dementias; ARIA-E = amyloid-related imaging abnormalities–edema/effusions; ARIA-H = amyloid-related imaging abnormalities–hemorrhage/hemosiderin deposition; ATT = amyloid-targeting therapy; CT = computed tomography; FFS = fee-for-service; ICD-10-CM = International Classification of Diseases, Tenth Revision, Clinical Modification; ICH = intracerebral hemorrhage; ID = identifier; MCI = mild cognitive impairment; MRI = magnetic resonance imaging; NLP = natural language processing.

Figure 1. Flow diagram of algorithms development and validation study.

8.2.1.4. Identification and Characterization of Potential Comparator Cohorts for Future Safety Assessment of Amyloid Targeting Therapies

To identify potential comparator cohorts for future safety assessment of ATTs, patient cohorts will be identified using diagnoses and medications from the de-identified PPD structured EMR data. A total of 2 potential comparator cohorts (Comparator Cohorts A and B) and an ATT reference cohort will be compiled and characterized. The goal of the characterization is to

explore suitable comparison groups for ICH, ARIA, seizures, anaphylaxis, and death outcomes in a future study.

Comparator Cohort A (disease cohort)

Comparator Cohort A inclusion criteria

- ADRD or MCI diagnosis during the study period from 01 January 2022 to the latest available data
    - Index event = first ADRD or MCI diagnosis during the study period
- Patients aged 65 years or older at index
- With at least 18 months of pre-index continuous enrollment (including the index day) in PPD data (where continuous enrollment is defined as at least 1 encounter during the 18 months prior to index)

Comparator Cohort A exclusion criteria

- Previous diagnosis of ADRD or MCI in the 18 months pre-index (not including the index day)
- Evidence of receiving any ATT (i.e., donanemab, lecanemab, and aducanumab; see Table 1) during the 18-month period pre-index (diagnosis of ADRD or MCI up to and including the index day)

Comparator Cohort B (AChEI cohort)

Comparator Cohort B inclusion criteria

- Acetylcholinesterase inhibitor (AChEI during the study period from 01 January 2022 to the latest available data
    - Index event = first prescription of an AChEI medications (i.e., donepezil, rivastigmine, and galantamine; see Table 2) between 01 January 2022 and the latest available data.
- Patients aged 65 years or older at index
- With at least 18 months of pre-index continuous enrollment (including the index day) in PPD data (where continuous enrollment is defined as at least 1 encounter during the 18 months prior to index)

Comparator Cohort B exclusion criteria

- Use of any AChEI (i.e., first prescription of donepezil, rivastigmine, galantamine; see Table 2) in the 18 months pre-index (not including the index day)
- Evidence of receiving any ATT during the 18 months pre-index (up to and including the index day)

Reference cohort

Under the assumption that donanemab-treated patients and patients receiving any other ATT (lecanemab or aducanumab) will share comparable characteristics in term of demographics, disease stage severity of illness, and comorbidities, and because of the recent launch of

donanemab that will limit the number of eligible donanemab-treated patients, the reference cohort will not be limited to donanemab-treated patients and will be composed of patients receiving any ATT (i.e., donanemab, lecanemab, and aducanumab; see Table 1). Additionally, patients receiving ATTs are likely to have symptomatic treatments such as AChEIs in their baseline, which prevents robust use of the incident new user design. Therefore, AChEIs will not be excluded from the reference cohort. It is likely that the prevalent new user design would need to be utilized in a future potential comparative analysis comparing ATTs to AChEIs.

Reference cohort inclusion criteria

- Evidence of receiving any ATT during the study period from 01 January 2022 to the latest available data
    - Index event = first ATT infusion during the study period
- Patients aged 65 years or older at index
- With at least 18 months of pre-index continuous enrollment (up to and including the index day) in PPD data (where continuous enrollment is defined as at least 1 encounter during the 18 months prior to index)

Reference cohort exclusion criteria

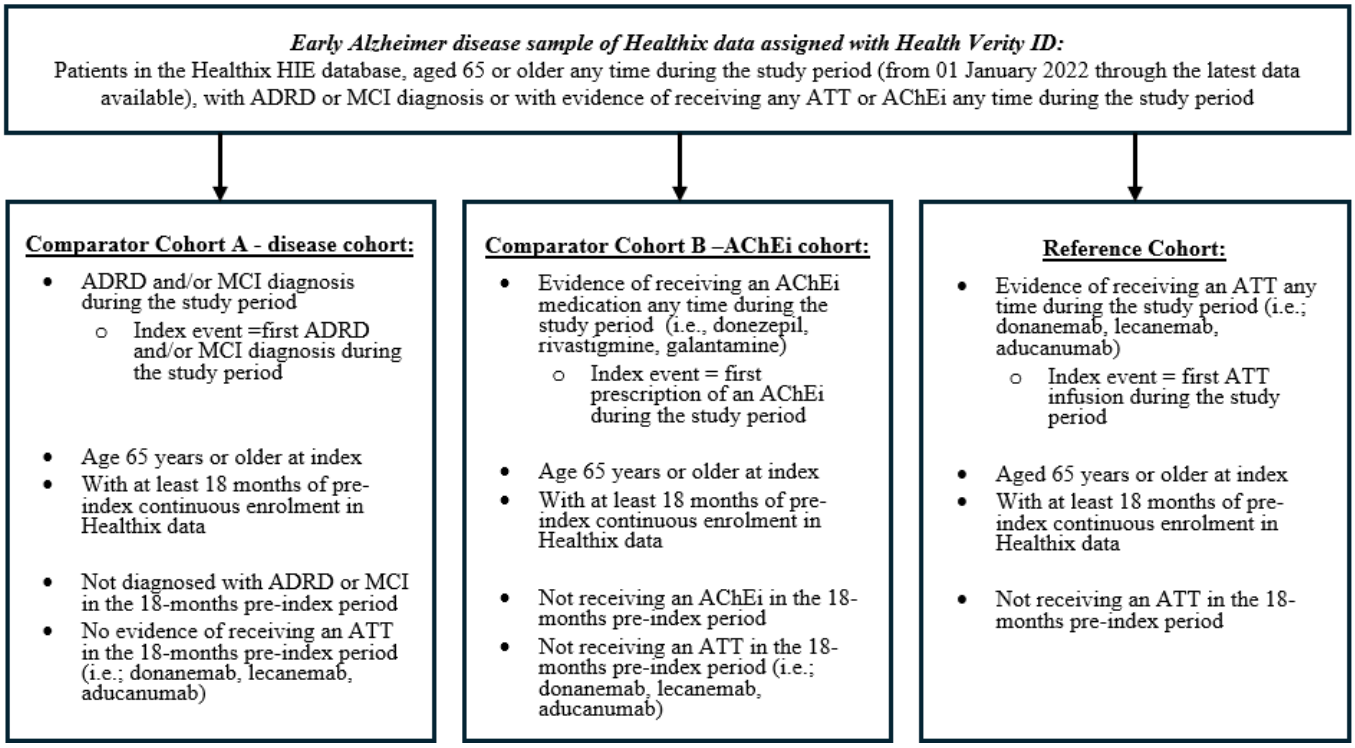
- Evidence of receiving any ATT during the 18-month pre-index event (not including the index day)

A descriptive summary for each patient-level cohort (Comparator Cohort A, Comparator Cohort B, and the Reference cohort) will be provided. No comparative analysis will be performed.

Table 2. Acetylcholinesterase Inhibitors

Nonproprietary Name	Proprietary Name
Donepezil, Donepezil Hydrochloride, Donepezil Base	Aricept, Adlarity
Rivastigmine, Rivastigmine Tartrate, Rivastigmine Transdermal System	Exelon
Galantamine, Galantamine Hydrobromide, Galantamine Benzoate Gluconate, Benzgalantamine	Zunveyl
Memantine and Donepezil, Memantine and Donepezil Hydrochlorides, Memantine Hydrochloride and Donepezil Hydrochloride Extended-release	Namzaric

The study populations used to identify potential comparator cohorts for future safety assessment are depicted in Figure 2.



Abbreviations: AChEI = acetylcholinesterase inhibitor; ADRD = Alzheimer's disease and related dementias; ATT = amyloid-targeting therapy; MCI = mild cognitive impairment.

Figure 2. Patient-level comparator and reference cohorts.

8.2.1.5. Identification of Published Algorithms for Future Safety Assessment of Seizure, Anaphylaxis, and Death

A review of recent literature will be conducted to describe EMR-based algorithms or alternative methods to identify seizure, anaphylaxis, and death in peer-reviewed publications. Applicable literature will be considered, and relevant articles will be described.

8.2.2. Rationale for Inclusion and Exclusion Criteria

Claims-based algorithms to identify ICH >1 cm and ARIA events are being developed in a separate protocol using US Medicare data, as part of a Kisunla FDA postmarketing requirement. In an effort to enable discussions regarding the performances of algorithms developed with different methods, comparable populations are targeted for inclusion.

As most patients with MCI or mild AD are aged 65 years or older, and to be consistent with the population selected for the Medicare claims analysis, patients aged 65 years and older are included in this study. Data since 2022 are included because the routine MRI monitoring of

ARIA events in ADRD and MCI population was not a common practice until after FDA approval of the first ATT in June 2021, and CMS covered ATTs since 09 January 2022. To be consistent with this time frame, the study will also assess ICH beginning in 2022.

According to a recent publication descriptively comparing characteristics of patients with ADRD who have Medicare Advantage versus Medicare FFS (Schroeder et al. 2024), substantial differences based on insurance type are not expected. Additionally, given the lag in Medicare Advantage data, the claims-based algorithms will be developed on Medicare FFS patients only.

The NLP algorithm to identify ICH >1 cm is developed with the aim of providing a reliable tool for assessing the safety outcome occurrence in a specific context of drug exposure. Nevertheless, as the risk of falls and subsequent traumatic injury is increased in Alzheimer's and cognitively impaired populations (Kornblith et al. 2024), it was decided to flag but not exclude TBI diagnoses, which can result in ICH >1 cm. The potential differences in imaging reports habits and the impact on the algorithm's performance are presumed to be minimal, though unpredictable and difficult to measure. Sensitivity analysis will be conducted and include reporting measures of diagnostic validity, indicators of performance, and measures of agreement separately for the subsample of patients with and without TBI, assessed in a window of 30-day look back and 7 days after the date of the qualifying imaging report.

Due to the relative rarity of the ICH event, even in the higher-risk AD population (Pinho et al. 2021), and to allow for a sufficient number of reference-standard positive cases for validation, evidence of a nontraumatic ICH determined via ICD-10-CM diagnosis in the PPD EMR structured data during the study period (Codes I61.x, See Standalone Document No. 3 for description) was included as an inclusion criteria for the Report Group 1 Validation Subgroup. This was done to increase the proportion of reference-standard positive cases ("True positive" and "False negative" cases) in the Report Group 1 Validation Subgroup, as these cases contribute in particular to sound estimation of algorithms sensitivity and PPV.

Also, the estimation of specificity and NPV required identification of a representative group of true-negative or algorithm-negative individuals (Ehrenstein et al. 2024). To this end, the Report Group 2 Validation Subgroup (composed of patients treated with ATT, recommended to receive routine monitoring MRIs, even if asymptomatic, see Section 8.2.1.3) will be combined with Report Group 1 Validation Subgroup (composed of patients with ADRD or MCI diagnoses, with at least 1 head CT or brain MRI, and with evidence of a nontraumatic ICH determined via ICD-10-CM diagnosis in the PPD EMR structured data, see Section 8.2.1.3), to evaluate the performance of the ICH >1 cm algorithm.

8.2.3. Patient Identification

There are no patients of special interest or subgroups to identify because the primary objective is to develop and estimate NLP algorithms performances.

8.3. Variables

The variables described in this subsection will be collected at the report level. As a result, each patient can contribute more than 1 data point if more than 1 imaging report meets the inclusion criteria. Separately, the comparator and reference cohorts will be identified at the patient level, and applicable variables will be assessed accordingly.

8.3.1. Drug Exposure or Study Treatment

Although some patients will have evidence of receiving ATT and/or AChEI medications, all data will be collected at the report level with no information on temporality between the imaging reports and treatment. Of note, a potential comparison between donanemab-treated patients and appropriate comparator groups might be performed in a future cohort study using secondary EMR data, which would be described in a separate study protocol.

8.3.2. Outcome Variables

The outcome variables will be identified using the developed NLP algorithms and the medical chart review (for validation subgroups only):

- Any ICH >1 cm: identified from head CT or brain MRI reports meeting inclusion criteria (categorical: yes = ICH >1 cm; no (no evidence) = no evidence of ICH; no (<1 cm) = ICH size <1 cm; no (size not provided) = ICH size not provided or unknown; See Section 8.7.2.3, Table 3 for examples)
- Any ARIA-E: identified from brain MRI reports meeting inclusion criteria (categorical: yes = edema or effusion; no (no evidence) = no evidence of edema or effusion; no (unknown) = information not provided or unknown)
- Any ARIA-H: identified from brain MRI reports meeting inclusion criteria
    - Any ARIA-H with superficial siderosis (categorical: yes = superficial siderosis; no (no evidence) = no evidence of superficial siderosis; no (unknown) = information not provided/unknown)
    - Any ARIA-H with microhemorrhage(s) (categorical: yes = microhemorrhage(s); no (no evidence) = no evidence of microhemorrhage(s); no (unknown) = information not provided or unknown)
    - ARIA-H overall will be constructed from the ARIA-H subtypes outcomes (i.e., any ARIA-H with superficial siderosis or any ARIA-H with microhemorrhage(s)) (dichotomous: yes= evidence of superficial siderosis OR evidence of microhemorrhage(s)); no = (no evidence of superficial siderosis OR unknown on superficial siderosis) AND (no evidence of microhemorrhage(s) OR unknown on microhemorrhage(s)). No specific algorithm will be developed for ARIA-H overall

8.3.3. Other Study Variables for Algorithms Development and Validation

Report-level data

Using structured PPD EMR data, the following demographics and general characteristics will be assessed at the report level, on the date of the imaging report (i.e., entry event) or the date of the nearest prior visit (if there is no date on the imaging report):

- age (continuous and categorical: i.e., 65-74, 75-84, and ≥85 years)
- sex (categorical: i.e., male, female, and unknown)
- race (categorical: i.e., White, Black, Asian, other, unknown)
- ethnicity (categorical: i.e., Hispanic or Latino, non-Hispanic or Latino, and unknown)
- primary insurance type (categorical: i.e., Medicare FFS, Medicare Advantage, other, and unknown), and
- type of imaging report (dichotomous: head CT and brain MRI).

Using structured data, the following characteristics will be assessed at the report level, allowing for a 1-year look-back period, up to and including the day of the imaging report (i.e., entry event):

- Comorbid conditions, as defined via ICD-10-CM diagnosis codes and descriptions, and as delineated in published work (Elixhauser et al. 1998; Quan et al. 2005; Rosenthal et al. 2017; van Walraven et al. 2009; Hsieh et al. 2021; Smith et al. 2020), where applicable (dichotomous for each condition: yes, no) (Standalone Document No. 3):
    - Comorbidities potentially influencing the propensity to get brain imaging: multiple sclerosis, epilepsy or seizure disorder, headache disorder, Parkison disease, any cancer (including brain tumor), delirium or psychosis
    - Other relevant medical brain history: ischemic stroke or transient ischemic attack, nontraumatic hemorrhagic stroke or intracerebral haemorrhage, other nontraumatic intracranial hemorrhage (subdural or epidural, traumatic ICH, other traumatic intracranial hemorrhage), meningoencephalitis or intracranial infection
    - Medical history of comorbidities frequently reported among the AD population: arterial hypertension, heart failure, diabetes, acute myocardial infarction or ischemic heart disease, mood disorders (anxiety/depression), peripheral vascular disease
- Evidence of antithrombotic medication (dichotomous: yes, no) (Standalone Document No. 2)

Using structured data, the following diagnosis codes will be assessed at the report level, allowing for a window of 30-day look back and 7 days after the date of the imaging report (i.e., entry event):

- Traumatic brain injury diagnosis (Warwick et al. 2020) (dichotomous: yes, no), which can include ICH >1 cm (ICD-10-CM diagnosis codes described in Standalone Document No. 3)
- Other brain pathology diagnosis codes which, if suspected (e.g., in presence of etiologically relevant symptoms), may have prompted the qualifying brain imaging procedure, including ischemic stroke or transient ischemic attack, nontraumatic hemorrhagic stroke or ICH, other nontraumatic intracranial hemorrhage (subdural or epidural), traumatic ICH, other traumatic intracranial hemorrhage, and meningoencephalitis or intracranial infection (dichotomous for each condition: yes, no) (ICD-10-CM diagnosis codes described in Standalone Document No. 3)

Patient-level data

The following information will also be assessed for the NLP algorithms to identify ICH >1 cm development and validation subsets:

- distribution of the number of CT imaging reports (categorical: e.g., 1, 2, 3, 4+) and the total number of CT reports (continuous) per patient
- distribution of the number of MRI imaging reports (categorical: e.g., 1, 2, 3, 4+) and the total number of MRI reports (continuous) per patient, and
- distribution of the number of CT and MRI imaging reports (categorical: e.g., 1, 2, 3, 4+) and the total number of CT and MRI reports (continuous) per patient.

The following information will also be assessed for the NLP algorithms to identify ARIA development and validation subsets:

- distribution of the number of MRI imaging reports (categorical: e.g., 1, 2, 3, 4+) and the total number of MRI reports (continuous) per patient

8.3.4. Other Study Variables for Description of Potential Comparator and Reference Cohorts

Using structured data, the following additional clinical characteristic will be assessed at the patient level, in the 18 months pre-index, including the index day:

- Evidence of ADRD or MCI, as described in Standalone Document No. 1 (dichotomous: yes, no)
- Evidence of AChEI medications, alone or in combination (donepezil, rivastigmine, and galantamine; as described in Table 2, Section 8.2.1.4, and Standalone Document No. 4) (dichotomous: yes, no)

- Evidence of memantine medication, alone or in combination (proprietary name, National Drug Codes, as described in Standalone Document No. 4) (dichotomous: yes, no)
- Evidence of ATT infusion (donanemab, lecanemab, or aducanumab; Table 1, Section 8.2.1.2) defined by the identification of generic or brand names, Healthcare Common Procedure Coding System codes, or National Drug Codes, as described in Standalone Document No. 4 (dichotomous: yes, no)

8.4. Data Sources

PPD HIE data

This study will use PPD HIE data. PPD securely exchanges data and serves in the New York downstate region, including New York City and Long Island, collecting data from more than 9000 healthcare facilities for more than 21 million patients. The data are aggregated across healthcare facilities from large healthcare systems to small community health centers and individual physician practices, including behavioral health and community-based organizations across the region. The trusted partner in sharing health information to improve people's lives, PPD is a core contributor and qualified entity of the Statewide Health Information Network of New York, which is the largest public HIE in the nation.

PPD will provide a subset of patient data that meet the inclusion criteria for this study, as will be described in the statistical analysis plan. These data will include de-identified structured data elements, such as demographics, diagnoses, procedures, and medications, and the unstructured data will be gleaned from the de-identified CT and MRI reports. PPD will run the HIE study data through PPD Identity Manager engine to produce a PPD ID for each study patient, and enable linkage for potential future studies.

The PPD data are taken directly from healthcare facilities' EMRs, with all records of patients treated in the facilities reported to the HIE. It includes patients treated in inpatient and outpatient settings, as the data sources include EMRs from hospitals and clinics, allowing for a representative sample of patients with ADRD and MCI. PPD will de-identify the data before delivering it for NLP and analysis. The data elements will be used to create study variables specified in Section 8.3.

8.4.1. Appropriateness of Data Source in Addressing Safety Questions of Interest

Development and validation of algorithms to identify imaging reports with ICH >1 cm, ARIA-E, and ARIA-H in a real-world data source will facilitate monitoring of these important ATT risks in large real-world populations.

To this end, this study will attempt to develop and validate NLP algorithms to detect the events of interest at scale: ICH >1 cm, ARIA-E, and ARIA-H in EMR records, specifically from unstructured radiology report narratives. NLP algorithms will be developed and validated for

future application with large real-world EMR data sets, allowing for a larger volume and faster identification of events than can be attained with manual chart review.

The PPD HIE data, a large, multifaceted data source, allows NLP algorithm development and validation to occur using a real-world population. PPD covers over 9000 healthcare facilities in the New York area, including hospitals, community health centers, and clinical practices. Given that the distribution of race and ethnicity is diversified in the New York area, minorities are well represented. In addition, with the inclusion of patients from hospitals and community health centers, underserved patients are also represented in the data. Using HIE data reduces the proportion of patients lost to follow-up because patients visiting different facilities or clinics can be tracked as long as they still seek care in the facilities covered in the HIE.

PPD provides clinical details from patients' EMRs and radiology reports for a large patient population treated in all clinical settings. The use of unstructured brain imaging reports using the NLP method provides a particularly relevant approach to identifying ICH >1 cm and ARIA events in secondary real-world data. Indeed, the ICD-10-CM does not contain any specific diagnosis codes for ICH >1 cm, ARIA-E, or ARIA-H, and imaging results are rarely entered in structured data. Therefore, a review of imaging reports will provide clinical details and retrieve relevant information. ICH is detected in head CT scans or brain MRI scans, and ARIA events are detected in brain MRI scans (most often in patients treated with ATTs).

The NLP algorithms will be developed and applied to the patients' head CT or brain MRI report(s) included in the HIE data to identify patients with ICH >1 cm and ARIA. The structured data, such as diagnosis and medication, will be used to identify eligible patient groups, categorize them accordingly, and describe the groups included in the study.

8.4.2. Enrollment and Comprehensive Capture of Care

Since this study involves secondary use of data, all patients meeting the patient selection criteria will be included in the study. The PPD HIE includes data from healthcare systems, community health centers, clinics, and physician practices, and therefore captures care in various settings.

8.4.3. Country of Origin

The PPD HIE is based in the US.

8.4.4. Selection of Study Population

The source population extracted from the PPD database to identify the study population included patients aged 65 years or older, with any diagnosis of ADRD or MCI, and/or evidence of treatment with ATTs or AChEIs that are reported in the PPD HIE data on or after 01 July 2020, which allows for up to an 18-month look-back period prior to the study start date of 01 January 2022. Patient reports will be assigned to an NLP algorithm development group and/or

validation subgroup, as applicable, from this study population. Patients will be assigned from this study population into the potential reference and comparator cohorts for future safety assessment of ATTs.

Inclusion and exclusion criteria for each study group are described in Section 8.2.1. As appropriate, patient-level and report-level criteria are outlined for NLP algorithm development of ICH and ARIA groups (Section 8.2.1.1and 8.2.1.2), algorithm validation of ICH and ARIA subgroups (Section 8.2.1.3), and potential reference and comparator cohorts (Section 8.2.1.4). To facilitate NLP algorithm development and validation, criteria are specified individually for each objective or study group.

Appropriateness of the Data Source for the study population is described in Section 8.4.1. The PPD HIE was selected as a real-world data source that could be used for development and validation of NLP algorithms of ICH >1 cm and ARIA in patients with ADRD and MCI because of the radiology reports availability, and for the de-identified structured and unstructured EMR. Using a large representative dataset supports Lilly's aim of assessing the frequency of these events in patients treated with donanemab in potential future safety studies.

3.1.1. Quality Assurance and Quality Control

PPD infrastructure is designed to support interoperability across multiple health care systems and is based on the InterSystems vendor platform (InterSystems, Cambridge, MA, USA) (Fleischman et al. 2014). Each participating healthcare organization has a dedicated local server that uses a common data model designed to support primary care use for patient-level encounters and clinical data. Each patient encounter is interfaced with the hub's master patient index using standard Health Level 7 messages (HL7 2024) and a probabilistic match is performed based on the patient demographic information and predetermined thresholds set to maintain a low false-positive match rate. This system operates in real-time to support primary clinical care and operates in parallel with a VDW that is used to support secondary-use cases. The VDW is organized using a relational database model built on a Microsoft SQL Server allowing analytics and reporting. Using specific Health Level 7 messages and data elements (i.e., patient medical record, encounter type, and admission and discharge dates) and master patient index ID, the patient can be tracked across multiple sites. Analytics and reporting across the entire HIE are facilitated because the VDW structure allows for near-real time updates and harmonizes data across multiple care settings. The VDW has undergone both technical and validation testing.

PPD performs quality checks on EMR data fed to the HIE on a regular basis. Premier will also perform a data quality check by examining missing values, invalid records, and outliers, and data cleaning will be performed accordingly.

Premier Healthcare Solutions (Premier) will work to build and train NLP algorithms to identify ICH >1 cm from head CT and brain MRI reports and ARIA-E and ARIA-H from brain MRI reports. After developing the NLP algorithms, a manual review of reports will be conducted to

estimate the validity and performance of the developed NLP algorithms. The manual review will involve 2 abstractors performing double, independent annotations, and a radiologist will reconcile discrepancies. Manual review of CT or MRI reports will follow guidelines outlined in the annotation (Section 8.7.2.3) and a separate Specification Document, which will be developed in collaboration with an SME. The performance measures of the developed NLP algorithms (e.g., sensitivity, specificity, PPV, NPV, accuracy, and F1 scores) in identifying ICH >1 cm, ARIA-E, ARIA-H with superficial siderosis, ARIA-H with microhemorrhage(s), and ARIA-H overall will be estimated, and the NLP algorithms will be adjusted accordingly.

8.4.5. Study Time Frame and Lag-Time Issues

The time frame of the current study will cover 01 January 2022 to latest data available at the time of data extraction. There is almost no lag from facilities to send records to the HIE, and approximately a 1-month lag for the data to be de-identified.

8.5. Study Size

A preceding feasibility assessment revealed that overall, PPD HIE includes 129,929 patients with a diagnosis of ADRD or MCI since 2016, with more than 2.5 million radiology documents. When patient selection criteria were narrowed to include only patients aged 65 years and older and a head CT or brain MRI report, feasibility counts of the PPD HIE data estimated 7336 unique patients annually. Because the study objectives are descriptive, all available applicable data will be considered, and a specific study size is not required to develop the algorithms.

For algorithm development of ICH >1 cm, sample size will be dependent on the number of patients with ADRD or MCI diagnoses in the database during the study period. The proportion of positive cases is directly influenced by the incidence of ICH >1 cm in the ADRD and MCI population. The incidence rate of ICH in patients with AD was reported at 3.4/1000 person-years (Pinho et al. 2021).

For algorithm development of ARIA-E and ARIA-H, the sample size will be dependent on the number of patients exposed to ATTs during the study period. The first ATT treatment (aducanumab) was launched in June 2021 in the US, followed by lecanemab and donanemab in January 2023 and July 2024, respectively. CMS provided reimbursement for ATTs starting January 2022. For algorithm development of ARIA-E and ARIA-H, the proportion of positive cases is directly influenced by the incidence of ARIA-E and ARIA-H in the ATT-treated population, which has been estimated at 24.4% and 31.4%, respectively (Lilly internal data).

Regarding sample size for validation of the algorithms using manual chart review, an article by Liu and colleagues (Liu et al. 2021) outlined a framework for calculating validation sample size by a priori identifying the critical lower confidence bound for the PPV/NPV ratio. In the following formula, $\mathbf{n}$ is the number of true positive cases needed, $\mathbf{z} = 1.96$ (critical value of the standard normal distribution with the Bonferroni correction), $\hat{\mathbf{p}}$ is the estimate of the PPV/NPV ratio and $\mathbf{p_0}$ is the critical lower confidence bound for the PPV/NPV ratio.

$$n = \frac{z^2_{\alpha/2}\hat{p}(1-\hat{p})}{(\hat{p}-p_0)^2}$$

In this case, as an identification of an algorithm for an imaging outcome such as ICH >1 cm, ARIA-E and ARIA-H may be expected to achieve lower PPV and higher NPV than other more straightforward algorithms, an estimated PPV/NPV ratio was set to 0.900 with a critical lower bound of 0.800. In pharmacoepidemiology studies, it's generally acknowledged that sample sizes should be sufficiently large to estimate the accuracy parameters, prioritizing validity measures (sensitivity and PPV) (Ehrenstein et al. 2024; Gillmeyer et al. 2021). Therefore, the Liu et al. (Liu et al. 2021) formula was utilized to calculate the target for true positive cases as follows:

$(1.96^2*0.9*0.1)/((0.9-0.8)^2)$ or 35 **true positive cases**.

For ICH >1 cm, given the low expected background incidence rate in the AD population (3.4 per 1000 person years) (Pinho et al. 2021), positive cases in the validation sample will be bolstered by requiring an ICD-10-CM diagnosis for nontraumatic ICH in PPD EMR structured data. However, the number of true positives for ICH >1 cm within this bolstered sample is still unknown. According to an assessment of ICH volume conducted in The Genetic and Environmental Risk Factor for Hemorrhagic Stroke study (Robinson et al. 2022), the median volume of ICH is 14 mL, which equates to ICH of 3 cm diameter (assuming sphere). As a consequence, it is expected that most diagnosed ICH will be greater than 1 cm. Assuming a conservative estimate of ICH greater than 1 cm of 25%, a sample size of 138 would be required to obtain 35 positive cases. Therefore, to account for the expectation of approximately 70% successful linkage between the charts and claims, a sample size of 198 will be utilized in Report Group 1 Validation Subgroup to ensure at least 35 positive cases are obtained.

For ARIA-E and ARIA-H, the expected number of positive cases can be obtained from the clinical trial results. For ARIA-E, with an expected incidence proportion of 24.4%, 35 cases is expected with a sample size of 142 patients. For ARIA-H, with an expected incidence proportion of 31.4%, 35 cases is expected with a sample size of 110. Therefore, to account for the expectation of approximately 70% successful linkage between the charts and claims, a sample size of 202 will be utilized in Report Group 2 Validation Subgroup to ensure at least 35 positive cases for ARIA-E and ARIA-H are obtained.

For the manual chart abstraction, all applicable imaging reports from the validation subgroups will be used to validate the ICH >1 cm and ARIA algorithms. Sample size for patients identified for the potential comparator cohorts and reference cohort for future safety assessment of amyloid targeting therapies will be determined based on the number of patients meeting inclusion and exclusion criteria for each cohort and are not based on a priori estimates.

8.6. Data Management

Patient data are recorded in data files and data forms embedded in the EMR, which in turn populates the HIE. Premier will use data files of the de-identified structured and unstructured HIE data. Premier research staff are responsible for the integrity of the data (i.e., accuracy, completeness, legibility, and timeliness) reported to Lilly.

8.6.1. Management of Data Used for Prediction Algorithm Development

This study will involve the secondary use of medical record data. PPD will de-identify the structured data (i.e., diagnoses, medications, and patient demographics) and unstructured data (i.e., head CT and brain MRI reports) in accordance with HIPAA requirements before delivering for analysis.

All data will be securely transferred and stored on a password-protected server accessible only to Premier's research staff. Analytic files and programs are retained as per Premier's "Information Classification and Retention Policy," which aligns with the National Institute of Standards and Technology security framework. Nonpublic information is kept in restricted directories, with administrative and physical safeguards compliant with state and federal laws and corporate policies. A virtual private network is used for secure remote access by necessary resources only. In accordance with these policies, appropriate data security controls, including data storage, data encryption, and password protection, are in place throughout the lifecycle of the research study to minimize risk of third-party data interception.

8.6.2. Management of Data for Selection of Reports Used for Algorithms Validation

A claims-based algorithm to identify ICH >1 cm and ARIA events is being developed in a separate protocol using Medicare US data, as part of a Kisunla FDA postmarketing requirement. In an effort to enable discussions regarding the algorithms performances using different methods (EMR NLP-based versus claims-based algorithms), the same validation subgroups will be utilized for both studies. To validate claims-based algorithms developed in this parallel project, outcomes data from PPD medical chart review must be linked to Medicare data. Linkage between the charts and Medicare claims will be facilitated by PPD through personal identifiable information available from a third party, PPD. A first identification of PPD study patients with a high likelihood of linkage with CMS data will be performed. The process flow is described below.

Assignment of PPD IDs to PPD data

PPD will statistically de-identify the structured data (i.e., diagnoses, medications, patient demographics) and unstructured data (i.e., head CT and brain MRI reports) in accordance with HIPAA requirements before delivering for analysis. PPD will install and host a version of

the PPD Identity Manager Batch De-ID Engine in a HIE secure environment to de-identify PPD PII study data and attribute PPD IDs to study IDs.

Identification of Medicare eligible patients using PPD ID

To enable linkage of Medicare FFS data with PPD data, through PPD tokens, this study will leverage an existing PPD Study ID match table. This match table is the result of a partnership between CareJourney by Arcadia (holder of a license to the CMS Virtual Research Data Center as a commercial innovator and PPD (a Social Determinant Of Health data firm) that was created after submitting a PPD National consumer file to CMS using a combination of first name, last name, date of birth, ZIP code. The list of de-identified PPD IDs for the PPD patients meeting the study inclusion and exclusion criteria, will be linked to Medicare FFS beneficiaries using the linked crosswalk table for PPD Study IDs.

PPD will provide the list of study PPD IDs that match to a PPD Study ID, for selection of the validation subgroups by Premier. To perform the manual review of imaging reports (head CT and brain MRIs), a total sample of 400 patients will be selected based on defined inclusion and exclusion criteria (Section 8.2.1.3) for a manual review of imaging reports, among PPD patients with a PPD Study ID and with FFS coverage as identified in EMR data.

The flow diagram of the validation study, including the selection of imaging reports for medical review, is depicted in Figure 1, Section 8.2.1.3.

8.6.3. Management of Data from Manual Review of CT/MRI Reports

A total sample of 400 patients will be randomly selected for a manual review of imaging reports. An eCRF (e.g., Microsoft Excel spreadsheet format), as described in the specification document, will be prepared and saved in a secure location. If no clinical findings are present in the report (e.g., “MRI performed at station 22”), then the other imaging reports for the same patient will be screened. If the patient does not have at least 1 imaging report with clinical finding, the manual review will not proceed, and the patient with missing findings will be excluded. Another patient will be randomly selected to replace the patient with missing data and manual review will proceed with their imaging report(s).

The final manual review data will be cleaned and merged to the tokens generated for use in the parallel claims-based algorithm study. Premier will deliver the data file to Lilly using a secure file transfer server.

8.6.4. Data Retention

Premier is bound by various obligations regarding the information it retains, the period of retention, and the process for destruction, which are outlined in its “Information Classification

and Retention Policy." Additional factors such as security classification, type of information, storage method, and business line are also relevant and covered in this policy.

At the conclusion of the study, data will be maintained for the minimum period that the longest applicable standard requires. At the end of this period, all study-specific data and digital records will be destroyed or securely deleted.

8.7. Data Analysis

This study will develop and refine a set of terms and logics (i.e., NLP algorithms) that can appropriately identify the presence of outcomes (ICH >1 cm, ARIA-E, and ARIA-H subtypes) from head CT and/or brain MRI reports. For the validation of the NLP algorithms, we will calculate the measures of diagnostic validity (i.e., sensitivity and specificity), indicators of performance (i.e., PPV and NPV), accuracy, and F1-scores. These statistical strategies are commonly employed in NLP methods (Ehrenstein et al. 2024; Fu et al. 2022; Verma et al. 2022; Zheng et al. 2024).. Descriptive analysis will be used to summarize patient demographic and clinical characteristics for imaging reports used in each algorithm subset, and to summarize the results from the algorithms validation. Characteristics of patients identified for the potential comparator and reference cohorts for future safety assessment of ATTs will also be summarized using descriptive analysis.

For assessment of ICH by the developed NLP algorithm, evidence of ICH and size >1 cm will be considered positive (i.e., indicated "Yes"), whereas evidence of ICH "ruled out" or size <1 cm will be considered negative with evidence (i.e., indicated as "no (no evidence)", "no, (size <1 cm)"). Reports where the information presented is not definitive (e.g., "size is larger than last report," "no change from prior image") will be categorized as an unknown negative (i.e., indicated as "No ICH size provided or unknown"). All negative reports (i.e., "no" categories: negative with evidence, unknown negative) will be combined into a single "No" category, which will be compared to reports with positive evidence of ICH >1 cm (i.e., "Yes" category). There will be no modification of raw data. Transformation of observed data points will occur in the context of assigning a category to free texts for each variable. For example, the free-text sentence "2.5 cm (AP) x 2.5 cm (TR)" associated to a hemorrhage will be transformed (or categorized) into "Yes" for the indicator of ICH >1 cm (Section 8.7.2.3, Table 3). Both unit size "cm" and "mm" will be considered (i.e., determination of ICH >1 cm = ICH >10 mm).

Similar methods of evaluating the MRI report narratives and categorization will be employed for ARIA-E and ARIA-H. All negative reports (i.e., "no" categories: negative with evidence, unknown negative) will be combined into a single "No" category, which will be compared to reports with positive evidence of the outcome (i.e., "Yes" category). Details for categorization of all outcomes will be delineated in the Specification Document and eCRF, which will be developed with input from an SME.

As the study will develop and validate NLP algorithms to identify real-world data on the occurrence of ICH >1 cm and ARIA, missing values will not be imputed. Missing values will be

categorized or set as "unknown" (if categorical), or missing (if numeric). Additionally, as no comparison is involved, there is no concern of confounding.

To allow for an adequately powered validation of the NLP algorithms, our strategy to increase the chances of having adequate numbers of positive cases is to sample patients with known diagnosis of ICH (in the case of ICH >1 cm), to sample patients receiving ATT (in the case of ARIA), and to use all available records for selected patients.

8.7.1. Analysis Overview

All study analyses will be descriptive in nature. Descriptive analysis will be used to summarize the performance of the NLP algorithms after validation and to characterize the report groups for algorithms development and algorithms validation and for the potential comparator cohorts created for the objectives.

Data measured on a continuous scale will be expressed as mean, standard deviation, median, interquartile range, minimum, and maximum. No tests of normality will be conducted. Categorical and dichotomous data will be expressed as counts and percentages of patients in the categories. Variable type (i.e., categorical, dichotomous, or continuous) is specified for each measure in Section 8.3 variable definitions.

Due to the HIPAA regulations, the counts and percentage for any table cell containing fewer than 11 patients will be masked and shown as "<11." As appropriate, variable categories may be combined to minimize the number of masked count table cells.

No comparative analysis will be done. All results will be descriptive summaries of the study population identified. All analyses will be performed using R (v.4.4.1 or higher).

8.7.2. Primary Analysis

8.7.2.1. Development of NLP Algorithms to Identify ICH > 1 cm and ARIA

Developing the NLP algorithms will include creating a set of terms and logics used to pull the outcome from relevant documents in the head CT and/or brain MRI reports. For each developed algorithm, refinements will be made through successive iterations of running the algorithm on the EMR data and re-checking the results (i.e., training set; see study populations for development of NLP algorithms to identify ICH >1 cm and ARIA, Sections 8.2.1.1 and 8.2.1.2). This method is an iterative process, and several rounds of refinements may occur until each algorithm can identify its respective outcome of interest with acceptable validity as determined by Premier's NLP team. Specifically, the Premier's research team, together with a trained NLP analyst and input from an SME (i.e., radiologist), will conduct refinements by manually reviewing a convenience sample of results (i.e., testing using the training set), then modifying the terms and logics of each algorithm accordingly. This process (i.e., refine, review, and test) will be repeated until the research team is satisfied with the general accuracy of the results.

8.7.2.2. NLP Algorithms Validation

To validate the NLP algorithms, a reference standard will be developed by 2 annotators via manual review of all qualifying radiology reports from the Report Group 1 Validation subgroup and the Report Group 2 Validation Subgroup (400 patients in total). Because Report Group 1 Validation Subgroup patients will have an ICD-10-CM diagnosis of nontraumatic ICH (Section 8.2.1.3) for whom many reports may have evidence of ICH >1 cm, also including reports from patients in Report Group 2 Validation Subgroup will ensure that some imaging reports will not have evidence of ICH >1 cm when validating the algorithm. On the other hand, since patients in Report Group 2 Validation Subgroup will have evidence of ATT (Section 8.2.1.3), which requires routine MRIs, including 1 prior to treatment initiation, an adequate number of positive and negative ARIA occurrence is anticipated, using imaging reports from 202 patients. For each outcome of interest (i.e., ICH >1 cm, ARIA-E, ARIA-H with superficial siderosis, and ARIA-H with microhemorrhages(s)), a categorical variable indicating the presence of the outcome will be determined for each report (i.e., for ICH >1 cm: yes/no (no evidence)/no (<1 cm)/no (size not provided); for ARIA-E and ARIA-H subtypes: yes/no (no evidence)/no (unknown); See Section 8.7.2.3). For the estimation of algorithms performances (i.e., validation), dichotomous variables will be created from the outcome variables described in Section 8.3.2: for each variable, all the "no" categories will be grouped and will contribute to negative cases according to the NLP algorithm (i.e., "True Negative" and "False Negative" cases in the confusion matrices).

Using an eCRF, 2 annotators will independently read each report (i.e., head CT and/or brain MRI scan report), and indicate the outcomes (i.e., double annotation). General details of this process are described in Section 8.7.2.3 and will be fully described in a separate Specification Document. A radiologist SME will review reports in which annotators disagree until a final standard is reached. The results of the final annotation will serve as the reference standard to estimate the validity and performance of the NLP algorithms.

Confusion matrices will be used in each instance of validation to calculate accuracy, sensitivity, specificity, PPV, NPV, for each outcome along with the F1-score positive (harmonic mean of sensitivity and PPV, also called recall and precision, respectively) and F1-score negative. The average values for measures of predictive accuracy and their 95% confidence interval will be reported following the validation:

PPV = True positive / (True positive + False positive) (equation 1)

NPV = True negative / (True negative + False negative) (equation 2)

Sensitivity = True positive / (True positive + False negative) (equation 3)

Specificity = True negative / (True negative + False positive) (equation 4)

Accuracy = (True positive + True negative) / (Total positive + Total negative) (equation 5)

F1-score positive = 2 × (PPV × Sensitivity) / (PPV + Sensitivity) (equation 6a)

F1-score negative = 2 × (NPV × Specificity) / (NPV + Specificity) (equation 6b)

For the outcome of ICH >1 cm, the NLP algorithm performance analysis will be evaluated overall (, using both CT and MRI imaging reports). Depending on sampling availability, algorithm performance may be stratified and evaluated only on CT and/or MRI reports.

The validity of using the NLP algorithm for ARIA-H with microhemorrhage(s) OR the NLP algorithm for ARIA-H with superficial siderosis to identify ARIA-H overall will be evaluated by calculating the performances measures previously described. The confusion matrix will combine cases of ARIA-H with microhemorrhage(s) or with superficial siderosis as follows:

True positive cases of ARIA-H overall = (True positive cases of ARIA-H with microhemorrhage(s) including True Positive cases of ARIA-H with superficial siderosis)

False positive cases of ARIA-H overall = (False positive cases of ARIA-H with microhemorrhage(s) including False Positive cases of ARIA-H with superficial siderosis)

True negative cases of ARIA-H overall = (True negative cases of ARIA-H with microhemorrhage(s) including True Negative cases of ARIA-H with superficial siderosis)

False negative cases of ARIA-H overall = (False negative cases of ARIA-H with microhemorrhage(s) including False Negative cases of ARIA-H with superficial siderosis)

A statistical analysis plan document will be created to describe the sampling strategy, validation method, and power for the validation of the NLP algorithms.

8.7.2.3. Annotation Specification

A Specification Document and eCRF will be developed with input from an SME (e.g., radiologist). Technical details regarding specific categorization of each outcome (i.e., ICH >1 cm, ARIA-E, and ARIA-H subtypes) will be determined after reviewing a sample of imaging report narratives. The SME and Premier team will ensure that guidelines for annotation are clear and comprehensive. A reference standard for each outcome will be defined so that categorization by annotators is consistent. Annotators will be trained using the Specification Document and eCRF on a sample of imaging reports.

The annotation will be recorded in an eCRF. Each annotator will have their own annotation eCRF (“individual annotation spreadsheet”), to ensure that they are blind to each other’s results. Once the annotation process is done, all individual annotation spreadsheets will be combined into 1 document (“annotation spreadsheet master”), which will contain columns indicating the patient ID, report ID, and annotator ID. This document will be analyzed for annotation disagreements. Outcomes in which annotators disagree will be reviewed by an SME, until alignment is reached. The results of the final annotation will serve as the reference standard for the validation of the NLP algorithms.

The individual annotation spreadsheet will contain patient ID, report ID, and 2 columns for each outcome (i.e., ICH >1 cm, ARIA-E, ARIA-H with superficial siderosis, and ARIA-H with microhemorrhage(s)). For each outcome, the first column will be in drop box format to indicate presence of outcome and will allow for the following answers: "Yes," "No evidence of event," and "No, information not provided/unknown." The outcome of ICH >1 cm will also include the answer option of "No, ICH <1 cm." The second column will be in free text format for the annotator to copy and paste the exact sentence found in the report that they used to base their decision in the previous column.

Note: The annotator should copy and paste the full sentence(s) so as to allow the SME to understand the annotator's decision and the context of the sentence without having to refer to the report.

Table 3 illustrates an example of an eCRF spreadsheet.

Table 3. Fictional Individual Annotation Spreadsheet

Patient_ID	Report_ID	Annotator	ICH >1 cm	ICH >1 cm_sentence*
1	1	1	Yes	MRI came positive for ICH; size:1.2 cm
1	2	1	No (size not provided)	Hemorrhage stable with prior
2	1	1	No (no evidence)	No hemorrhage found in exam
3	1	1	No (<1 cm)	Microhemorrhage; size 0.4 cm

*Example sentences that may be found in the radiology report.

8.7.2.4. Characterization of Report Groups Used for Algorithms Development and Algorithms Validation

Descriptive analysis will be used to summarize characteristics outlined in Section 8.3. Demographic and clinical characteristics will be summarized for each of the NLP algorithms development and algorithms validation sample of reports, including:

- NLP algorithm to identify ICH >1 cm development sample (Report Group 1)
- NLP algorithms to identify ARIA-E and/or ARIA-H subtypes development sample (Report Group 2)
- NLP algorithm to identify ICH >1 cm validation sample (Report Group 1 Validation Subgroup and Report Group 2 Validation Subgroup, total 400 patients), and
- NLP algorithms to identify ARIA-E and/or ARIA-H subtypes validation sample (Report Group 2 Validation Subgroup, 202 patients).

Since algorithm development and validation will occur at the report level, each patient may contribute more than 1 data point and characteristics of a patient may be reflected in more than 1 descriptive summary. The number of patients contributing reports and a distribution of number of reports contributed by patients in each of the groups will be reported. Frequency (count and

percentage) and distribution of the type of imaging report (i.e., CT and MRI) will be summarized for each sample as applicable. No comparative analysis will be performed in this study.

8.7.2.5. Description of the Potential Comparator Cohorts for Future Assessment of Amyloid Targeting Therapies

Patients who meet criteria (as described in Section 8.2.1.4) for the potential comparator cohorts for future safety assessment of ATT will be identified. For the 2 comparator cohorts and the reference cohort, descriptive statistics will be used to summarize patient demographics, general characteristics, comorbid conditions, and clinical characteristics (Section 8.3) at the patient level. Given inclusion/exclusion criteria for these cohorts (Section 8.2.1.4), a patient may be described in more than 1 cohort. No comparative analysis will be performed in this study.

8.7.2.6. Identification of Published Algorithms for Future Safety Assessment of Seizure, Anaphylaxis, and Death

Published EMR-based algorithms or alternative methods (e.g., linkage to mortality registry) to identify seizures, anaphylaxis, and death will be reviewed and described in the final study report. Algorithms performances from published validation studies will be reported as applicable. No statistical analysis will be performed.

8.7.3. Secondary Analysis

None.

8.7.4. Sensitivity Analysis

The performance of each developed NLP algorithm will be assessed considering diagnosis of TBI. Additionally, the performance of each developed NLP algorithm will be assessed considering diagnosis of other brain pathology if this diagnostic group is present at a substantial frequency. Sensitivity analysis will include reporting measures of diagnostic validity, indicators of performance, and measures of agreement separately for the subsample of patients with and without TBI and potentially for the subsample of patients with and without other brain pathology. No additional algorithm(s) will be developed separately for patients with and without TBI or patients with and without other brain pathology.

8.8. Quality Control

In total, 2 experienced principal research scientists will conduct the study, including secondary data collection, methods, analysis, result interpretation, and reports; a senior principal at Premier will oversee the work, with input and feedback from Lilly's research team. A biostatistician at Premier will review and approve all statistical methods. A primary senior research analyst will conduct NLP search and programming for this project, and a separate senior researcher analyst (validation analyst) will validate it. For all data processing steps, the validation analyst will review the programs along with input and outpatient datasets. For the analysis steps of the

project, double programming techniques to reduce the potential for programming errors will be employed.

Data in the HIE undergoes both technical and validation testing as part of the VDW that is used to support secondary-use analysis. On a regular basis, PPD conducts quality checks on the EMR data fed to the HIE. Premier will also evaluate the data for quality and validity.

8.9. Limitations of the Research Methods

As this is a secondary data use study using observational data, it is subject to a few limitations. First, the details provided in the radiology reports may vary by healthcare providers and by the type of brain imaging performed (e.g., CT versus MRI and MRI sequences used). Some reports may include detailed features, and some may only provide a general description, with or without suspected diagnosis. For example, some radiology reports may specifically mention "ARIA-H superficial siderosis," but others may only mention an "amyloid-related abnormalities," or provide a description of observed lesions without specifying the suspected diagnosis (for e.g., "small leptomeningeal haemosiderin deposit in the left frontal lobe sulcus"). This study will design the ARIA-H algorithm to capture the phenotypes of ARIA-H with microhemorrhage(s) and ARIA-H with superficial siderosis.

Second, given the recent launch of ATTs as a therapeutic option, the sample sizes of patients presenting ARIA events may be limited. Additionally, due to the relative rarity of ICH in the AD population (Pinho et al. 2021), the sample size of patients identified with ICH size >1 cm may also be limited. To address these challenges, the research team will identify 2 report groups (Report Groups 1 and 2) to develop an algorithm for identifying ICH >1 cm and algorithms to identify ARIA events, respectively. Specifically, to allow for the algorithms to have sufficient sensitivity to detect those events, Report Group 2 will include patients who have received an ATT and are thus more likely to experience an ARIA event.

Regarding algorithms validation, an independent selection of the population for validation allows estimation of sensitivity, specificity, PPV, and NPV (Ehrenstein et al. 2024). In this study, and for feasibility reasons described Section 8.2.2), the ICH >1 cm algorithm validation subset will be composed in majority of patients with evidence of a nontraumatic ICH (determined via I61.x ICD-10-CM diagnosis; maximum 198 patients from Report Group 1 Validation Subgroup; See Section 8.2.1.3). This was intended to increase the proportion of reference-standard positive cases in Report Group 1 Validation Subgroup ("True positive" and "False Negative" cases) as these cases contribute in particular to sound estimation of algorithms sensitivity and PPV.

As a consequence of this methodological choice, reference-standard negative cases may be underrepresented (i.e., "True negative" and "False positive" cases), and affect the estimation of algorithms performances. In case of low numbers of reference-standard negative cases (i.e., "True negative" and "False positive" cases), the PPV could be overestimated by the impact on the denominator. Nevertheless, the addition of the Report Group 2 Validation Subgroup to validate

ICH algorithm (202 patients treated with ATT) will mitigate this risk, and this methodologic choice will preserve the estimation of algorithm PPV.

Finally, non-hospital-related death is not well captured in EMR. Thus, alternative methods that could be used to identify death for a potential future safety assessment of this outcome posttreatment with amyloid targeting therapies will be described.

8.10. Other Aspects

Not applicable.

9. Protection of Human Subjects

Observational studies will be submitted to ethical review boards for approval as required by local law.

This study will be conducted in accordance with applicable laws and regulations in the US, where the study is being conducted.

This is an observational study using secondary data of PPD HIE. All data will be de-identified prior to study inclusion and assessment and as such is not considered human subjects research. Study data and recorded information cannot be identified directly or through identifiers linked to individuals. All data are compliant with HIPAA regulations. As a result of these factors and US federal regulation 45 Code of Federal Regulations 46, the study will seek exemption from institutional review board evaluation and informed consent.

10. Management and Reporting of Adverse Events/Adverse Reactions

10.1. Secondary Data Use Study

This is a noninterventional study based on secondary data use, and therefore, no individual case safety report reporting is required. This study has no protocol-defined AEs, so a summary of AEs cannot be included in the final study report.

10.2. Product Complaints

Lilly collects product complaints on marketed Lilly products, such as drugs, drug/device combinations, medical devices, software as medical device (e.g., mobile medical applications), and comparator product(s) used in postmarketing medical research studies to ensure the safety of study participants, to monitor quality, and to facilitate process and product improvements.

For Lilly products under evaluation and/or Lilly products not under evaluation but discovered in the course of the study, study personnel are instructed to report product complaints as they would for products in the marketplace.

For non-Lilly products, such as comparator drugs or medical devices, or concomitant drugs or medical devices, study personnel are instructed to report product complaints as they would for products in the marketplace.

11. Plans for Disseminating and Communicating Study Results

Final reports are not required to be submitted to regulatory agencies. The study, including the final report, will be registered in the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance Registry. The study findings may be submitted to a scientific congress and/or to a peer-reviewed journal.

12. References

Alzheimer's Association. Alzheimer's Disease Facts and Figures. (2024). https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf

Cordonnier C, Leys D, Dumont F, et al. What are the causes of pre-existing dementia in patients with intracerebral haemorrhages? *Brain*. 2010;133(11):3281-3289. https://doi.org/10.1093/brain/awq246

Ehrenstein V, Hellfritzsch M, Kahlert J, et al. Validation of algorithms in studies based on routinely collected health data: general principles. *Am J Epidemiol*. 2024;193(11):1612-1624. https://doi.org/10.1093/aje/kwae071

Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27. https://doi.org/10.1097/00005650-199801000-00004

Fleischman W, Lowry T, Shapiro J. The visit-data warehouse: enabling novel secondary use of health information exchange data. *eGEMS (Wash DC)*. 2014;2(1):1099. https://doi.org/10.13063/2327-9214.1099

Fu S, Lopes GS, Pagali SR, et al. Ascertainment of delirium status using natural language processing from electronic health records. *J Gerontol A Biol Sci Med Sci*. 2022;77(3):524-530. https://doi.org/10.1093/gerona/glaa275

Gillmeyer KR, Nunez ER, Rinne ST, et al. Development and validation of algorithms to identify pulmonary arterial hypertension in administrative data. *Chest*. 2021;159(5):1986-1994. https://doi.org/10.1016/j.chest.2020.12.010

Hampel H, Hardy J, Blennow K, et al. The amyloid-beta pathway in Alzheimer's disease. *Mol Psychiatry*. 2021;6(10):5481-5503. https://doi.org/10.1038/s41380-021-01249-0

Hampel H, Elhage A, Cho M, et al. Amyloid-related imaging abnormalities (ARIA): radiological, biological and clinical characteristics. *Brain*. 2023;146(11):4414-4424. https://doi.org/10.1093/brain/awad188

HL7. Health Level Seven International (HL7) website. Accessed June 23, 2025. https://www.hl7.org

Hsieh MT, Huang KC, Hsieh CY, et al. Validation of ICD-10-CM Diagnosis Codes for Identification of Patients with Acute Hemorrhagic Stroke in a National Health Insurance Claims Database. *Clin Epidemiol*. 2021;13:43-51. https://doi.org/10.2147/CLEP.S288518

Kidwell CS, Wintermark M. Imaging of intracranial haemorrhage. *Lancet Neurol*. 2008;7(3):256-267. https://doi.org/10.1016/S1474-4422(08)70041-3

Kisunla [prescribing information]. Indianapolis, IN: Eli Lilly and Company; 2024.

Kornblith E, Diaz-Ramirez LG, Yaffe K, et al. Incidence of traumatic brain injury in a longitudinal cohort of older adults. *JAMA Netw Open*. 2024;7(5):e2414223. https://doi.org/10.1001/jamanetworkopen.2024.14223

Liu L, Bustamante R, Earles A, et al. A strategy for validation of variables derived from large-scale electronic health record data. *J Biomed Inform*. 2021;121:103879. https://doi.org/10.1016/j.jbi.2021.103879

LoPresti MA, Bruce SS, Camacho E, et al. Hematoma volume as the major determinant of outcomes after intracerebral hemorrhage. *J Neurol Sci*. 2014;345(1-2):3-7. https://doi.org/10.1016/j.jns.2014.06.057

Ma C, Hong F, Yang S. Amyloidosis in Alzheimer's disease: pathogeny, etiology, and related therapeutic directions. *Molecules*. 2022;27(4):1210. https://doi.org/10.3390/molecules27041210

Manly JJ, Jones RN, Langa KM, et al. Estimating the prevalence of dementia and mild cognitive impairment in the US: the 2016 Health and Retirement Study Harmonized Cognitive Assessment Protocol Project. *JAMA Neurol*. 2022;79(12):1242-1249. https://doi.org/10.1001/jamaneurol.2022.3543

Petersen RC, Aisen P, Boeve BF, et al. Mild cognitive impairment due to Alzheimer disease in the community. *Ann Neurol*. 2013;74(2):199-208. https://doi.org/10.1002/ana.23931

Petersen RC, Lopez O, Armstrong MJ, et al. Practice guideline update summary: mild cognitive impairment: report of the guideline development, dissemination, and implementation subcommittee of the American Academy of Neurology. *Neurology*. 2018;90(3):126-135. https://doi.org/10.1212/WNL.0000000000004826

Pinho J, Quintas-Neves M, Dogan I, et al. Incident stroke in patients with Alzheimer's disease: systematic review and meta-analysis. *Sci Rep*. 2021;11(1):16385. https://doi.org/10.1038/s41598-021-95821-x

Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130-1139. https://doi.org/10.1097/01.mlr.0000182534.19832.83

Rabinovici GD, Gatsonis C, Apgar C, et al. Association of amyloid positron emission tomography with subsequent change in clinical management among medicare beneficiaries with mild cognitive impairment or dementia. *JAMA*. 2019;321(13):1286-1294. https://doi.org/10.1001/jama.2019.2000

Rajan KB, Weuve J, Barnes LL, et al. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). *Alzheimers Dement*. 2021;17(12):1966-1975. https://doi.org/10.1002/alz.12362

Robinson D, Van Sanford C, Kwon SY, et al. What is the median volume of intracerebral hemorrhage and is it changing? *Int J Stroke*. 2022;17(5):576-582. https://doi.org/10.1177/17474930211032594

Rosenthal NA, Chung J, Gundrum J, et al. Updated Coding Algorithm for Assessing Charlson Comorbidity Index using Large Hospital Administrative Data. Abstract (or Poster) presented at: ISPOR 22nd Annual International Meeting; May 2017. Boston, MA.

Rost NS, Smith EE, Chang Y, et al. Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. *Stroke*. 2008;39(8):2304-2309. https://doi.org/10.1161/STROKEAHA.107.512202

Schroeder KM, Afonso AS, Wang H, et al. Demographics, comorbidities, and comedications in newly diagnosed patients with Alzheimer's disease and related dementias: findings from United States Medicare claims data. *J Alzheimers Dis*. 2024;102(2):JAD231488. https://doi.org/10.3233/JAD-231488

Sims JR, Zimmer JA, Evans CD, et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*. 2023;330(6):512-527. https://doi.org/10.1001/jama.2023.13239

Smith JR, Jones FJS, Fureman BE, et al. Accuracy of ICD-10-CM claims-based definitions for epilepsy and seizure type. *Epilepsy Res*. 2020;166:106414. https://doi.org/10.1016/j.eplepsyres.2020.106414

US-DHSS. About Underlying Cause of Death, 1999-2020. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics. Accessed April 2024. https://wonder.cdc. gov/ucd-icd10.html

van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's Disease. *N Engl J Med*. 2023;388(1):9-21. https://doi.org/10.1056/NEJMoa2212948

van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care*. 2009;47(6):626-633. https://doi.org/10.1097/MLR.0b013e31819432e5

Verma AA, Masoom H, Pou-Prom C, et al. Developing and validating natural language processing algorithms for radiology reports compared to ICD-10 codes for identifying venous thromboembolism in hospitalized medical patients. *Thromb Res*. 2022;209:51-58. https://doi.org/10.1016/j.thromres.2021.11.020

Warwick J, Slavova S, Bush J, Costich J. Validation of ICD-10-CM surveillance codes for traumatic brain injury inpatient hospitalizations. *Brain Inj*. 2020;34(13-14):1763-1770. https://doi.org/10.1080/02699052.2020.1849801

Xiong L, Reijmer YD, Charidimou A, et al. Intracerebral hemorrhage and cognitive impairment. *Biochim Biophys Acta*. 2016;1862(5):939-944. https://doi.org/10.1016/j.bbadis.2015.12.011

Zheng C, Ackerson B, Qiu S, et al. Natural language processing versus diagnosis code-based methods for postherpetic neuralgia identification: algorithm development and validation. *JMIR Med Inform*. 2024;12:e57949. https://doi.org/10.2196/57949

Zimmer JA, Ardayfio P, Wang H, et al. Amyloid-related imaging abnormalities with donanemab in early symptomatic Alzheimer disease: secondary analysis of the TRAILBLAZER-ALZ and ALZ 2 randomized clinical trials. *JAMA Neurol*. 2025;82(5):461-469. https://doi.org/10.1001/jamaneurol.2025.0065

Annex 1. List of Standalone Documents

No.	Document Reference No.	Title
1	VV-PVG-121517	List of ADRD and MCI Diagnoses
2	VV-PVG-121518	List of Antithrombotic Agents
3	VV-PVG-121519	List of Diagnoses
4	VV-PVG-121520	List of Other Medications
5	VV-PVG-123162	Responsible parties

Abbreviation: No. = number.

Annex 2. Additional Information

Expertise and credentials of the study team

The multidisciplinary team consists of experts specialized in outcomes research, radiology, real-world evidence research, data science, and NLP of unstructured radiology reports, who will cover all aspects needed to ensure the success of the study. Study personnel have advanced, doctoral, and medical degrees in areas such as epidemiology, statistics, health economics, analytics, brain imaging, and biomedical engineering. The team also has clinical and research knowledge in ADRD and MCI.

The investigators have extensive experience in analyzing different data sources, including applying NLP algorithms to extract attributes from unstructured data, accessing clinical outcomes via manual chart review, and examining and summarizing structured real-world data from large databases. The team is proficient in combining de-identified data and curating datasets. Results of prior research have been used to support a range of regulatory needs, FDA submissions, and therapeutic areas. In previous studies using PPD data, the team has used NLP to identify clinical conditions, including pulmonary nodules size, number, and locations from unstructured radiology reports.