ELSEVIER

Alzheimer's & Dementia ■ (2015) 1-10

Alzheimer's & Dementia

Research Article

Development of a unified clinical trial database for Alzheimer's disease

Jon Neville[a], Steve Kopko[b], Steve Broadbent[a], Enrique Avilés[a], Robert Stafford[a], Christine M. Solinsky[c], Lisa J. Bain[d], Martin Cisneroz[a], Klaus Romero[a], Diane Stephenson[a,*], for the Coalition Against Major Diseases

[a]*Coalition Against Major Diseases (CAMD), Critical Path Institute, Tucson, AZ, USA*
[b]*CDISC, Austin, TX, USA*
[c]*USC Department of Pharmaceutical Sciences, USA*
[d]*Independent Science Writer, Elverson, PA, USA*

Abstract

Background: Data obtained in completed Alzheimer's disease (AD) clinical trials can inform decision making for future trials. Recognizing the importance of sharing these data, the Coalition Against Major Diseases created an Online Data Repository for AD (CODR-AD) with the aim of supporting accelerated drug development.

Objective: The aim was to build an open access, standardized database from control arm data collected across many clinical trials.

Methods: Comprehensive AD-specific data standards were developed to enable the pooling of data from different sources. Nine member organizations contributed patient-level data from 24 clinical trials of AD treatments.

Results: CODR-AD consists of control arm pooled and standardized data from 24 trials currently numbered at 6500 subjects; Alzheimer's Disease Assessment Scale-cognitive subscale 11 is the main outcome and specific covariates are also included.

Conclusions: CODR-AD represents a unique integrated standardized clinical trials database available to qualified researchers. The pooling of data across studies facilitates a more comprehensive understanding of disease heterogeneity.

© 2015 The Alzheimer's Association. Published by Elsevier Inc. All rights reserved.

Keywords: Alzheimer's disease; Clinical trials database; Placebo data; Data standardization; Data integration; Facilitated access

1. Introduction

Alzheimer's disease (AD) currently affects more than 36 million people worldwide, with the prevalence expected to triple by 2050 [1]. Yet, despite intensive efforts, there are no approved disease-modifying products capable of slowing or arresting the disease. Recent trials of AD drugs have raised concerns about the path forward for drug development and highlighted the importance of learning as much as possible from trials that have already been conducted for therapeutic candidates. Sharing the data collected in those trials has thus been recognized as an essential, albeit challenging, component of drug development efforts [2].

The U.S. Food and Drug Administration (FDA) recognizing the urgency of addressing the public health crisis that stems from a failure to translate scientific progress into new therapies, launched the Critical Path Initiative in 2004 [3] to the drive innovation for the treatment of major diseases such as AD, cancer, and diabetes. In 2005, Critical Path Institute (C-Path) was created as a public–private partnership to deliver on the mission of the Critical Path Initiative, specifically to improve the efficiency of drug and medical device development through the creation of broadly

*Corresponding author. Tel.: +1-520-547-3440; Fax: +1-520-547-3456.

E-mail address: Dstephenson@c-path.org

http://dx.doi.org/10.1016/j.jalz.2014.11.005
1552-5260/© 2015 The Alzheimer's Association. Published by Elsevier Inc. All rights reserved.

accepted standards and tools. C-Path is a fully independent, 501(c)3 nonprofit institute comprised of seven precompetitive consortia (www.cpath.org), including the Coalition Against Major Diseases (CAMD). The mission of CAMD is to develop new technologies and methods to accelerate progress in treating neurodegenerative diseases, namely AD and Parkinson's disease. CAMD serves as a neutral third party and brings together pharmaceutical companies, research organizations, patient advocacy organizations, regulatory and other government agencies, and academia to address critical needs in three major cross-cutting areas: data sharing, disease modeling, and biomarkers [4,5].

Among the first issues addressed by CAMD was the need to combine disparate clinical data contributed by multiple organizations. The Alzheimer's Disease Neuroimaging Initiative (ADNI) [6] provides an instructive example of how data sharing fuels progress. However, as ADNI is purely observational, there is a need to understand how the analysis of disease progression in ADNI subjects compares to that observed in other populations, particularly clinical trial subjects enrolled at multiple global sites. Therefore, it is essential to obtain data from randomized samples of subjects that are more representative of global clinical trial populations.

This manuscript describes the process by which CAMD developed an online repository for clinical trial data obtained in globally executed randomized controlled AD clinical studies (C-Path Online Data Repository-Alzheimer's disease; CODR-AD).

2. Methods

2.1. The CDISC standard, study data tabulation model

Establishing and conforming to comprehensive data standards was essential to the development of a database that enables the pooling of data from different sources. For this, CAMD partnered with the Clinical Data Interchange Standards Consortium (CDISC) [7], a nonprofit organization that focuses on developing global standards for clinical trial data collection. CDISC standards are preferred by regulators, industry, and other research organizations as a means of facilitating regulatory review, aggregation, and querying of data, sharing data between entities, and streamlining the acquisition and analysis of data. In 2012, when the Prescription Drug Free User Act was reauthorized, CDISC was recognized as an example of an organization that develops the kind of open standards needed for ensuring efficient review of medical products—standards that will be required for regulatory submissions to the agency by the end of FY2017 [8].

The foundational Study Data Tabulation Model (SDTM) standard as it existed at the start of CODR development was insufficient with regard to representing the AD-specific data of interest to CAMD. To address this issue, CAMD worked with CDISC to develop a previously nonexistent AD therapeutic area standard to accommodate additional data elements relevant to AD clinical trials. This therapeutic area standard included scores from the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) and the Mini-Mental State Examination (MMSE), β-amyloid, and tau biomarkers, and apolipoprotein E (*APOE*) genotype because the presence of the *APOE* ε4 allele is the strongest genetic risk factor for AD thus far identified [9].

Because a key goal of the database was to support the development of quantitative modeling and simulation tools, the variables and domains selected for standardization were those deemed necessary for developing a drug-disease-trial model [10]. The proposed AD-specific standards, developed by a team of clinical trial researchers and data standards experts, were reviewed and vetted through a public review and comment process. The resulting standards for AD clinical trials were published [11], representing the first disease-specific therapeutic standards. A summary of the more salient concepts captured by SDTM domains contained in the database is provided in Table 1.

As development of the standards progressed, it became increasingly clear that the standards would—in addition to facilitating the pooling of data from legacy clinical trials—also provide a resource for prospectively collecting data in new trials without the need for remapping after the fact.

SDTM defines how clinical study data should be structured for submissions to the FDA and other regulatory authorities. SDTM is suited for collecting data of various types and storing it in a relatively small number of observation classes. For example, it allows the preservation of all data collected at an individual visit by making use of "long" data structures. "Long" data sets are generally preferred over "wide" data sets for storing data when subject measures are repeated longitudinally. In a long data set, the variable itself is a column heading and separate observations are captured in different rows. In contrast, in a "wide" data set, each observation is captured as a separate variable (i.e., in a separate column). Long structures thus lead to fewer "holes" in the data set when some subjects have more observations than others, or when some subjects are missing some of the observations. Long data sets also facilitate the development of standardized programs to operate on this fixed standard data format. Conversely, wide data sets are generally more preferred for data capture and some types of analysis. Although the long database structure may be less intuitive to researchers accustomed to working with analysis subsets, the flexibility was important because the AD database includes disparate data and heterogeneous subjects. Thus, SDTM was appropriate for the intended CAMD database, given the longitudinal measures repeated across time in AD trials, particularly when the number of observations varies between subjects. Transforming between the two formats is typically a simple task in most statistical software packages.

2.2. Collecting and standardizing data

With the standards in place, patient-level data from the control arms of relevant trials were remapped and used to populate the database. The scope of patient-level data

Table 1
SDTM domains used for CODR-AD

CDISC domain	Abbreviation	Observation class	Contents
Demography	DM	Special purpose	Age Gender Race Ethnicity Country
Subject characteristics	SC	Findings	*APOE* genotype *MTHFR genotype
Concomitant medications	CM	Interventions	†Acetylcholinesterase inhibitors †Memantine †General medications
Adverse events	AE	Events	event Severity Duration
Medical history	MH	Events	primary diagnosis (MCI or AD) Family history of AD General medical history
Vital signs	VS	Findings	SBP, DB Heart rate Temperature Weight, height BMI Respiratory rate
Questionnaires	QS	Findings	ADAS-Cog MMSE Others as collected may be present, but not standardized
Laboratory results	LB	Findings	All labs collected, mapped to SDTM. Controlled terminology compliance was out of scope in LB.

Abbreviations: CODR-AD, Coalition Against Major Diseases created an Online Data Repository for AD; MCI, mild cognitive impairment; AD, Alzheimer's disease; SBP, systolic blood pressure; DBP, diastolic blood pressure; CDISC, Clinical Data Interchange Standards Consortium; BMI, body mass index; ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive subscale; MMSE, Mini-Mental Health Examination.

NOTE. A list of CDISC Study Data Tabulation Model domains (data sets) used in CODR-AD, their corresponding observation classes, and a summary of the types of data stored in each.

*Not all studies collected genotype; some study sponsors collected but did not provide this information due to issues with informed consent.

†Contributors were asked to supply memantine and acetylcholinesterase inhibitors names as generic when present in the data. All other concomitant medication names are provided verbatim (as collected), and may or may not be decoded to generic names.

requested was extensive, including characteristics such as demographics, medical history, subject disposition, and outcomes data such as longitudinal ADAS Cog and baseline MMSE scores, and adverse events. The common denominator longitudinal cognitive endpoint was ADAS Cog 11, which at the case-report-form level, showed different item administration orders between sponsors. A concordance analysis was performed, which showed that despite this item administration variation, the level of agreement was adequate and should not affect the interpretation of the ADAS Cog 11 (unpublished data). Additionally, members were asked to contribute tables of subject-visit records, disposition events, laboratory results, vital signs, and concomitant medications. *APOE* genotype was requested in the form of categorical, allele-level isoform data when available (each of two alleles could have a categorical value of 2, 3, or 4, enabling the derivation of a comprehensive *APOE* genotype by patient). CAMD partnered with the data standards experts within each member company to remap existing clinical data to the newly developed AD CDISC standard. This resource-intensive step was a critical success factor for the consortium, as data were disparate in nature across studies and could not be pooled for analysis in their original form.

The process of remapping data was accomplished in two stages: logical mapping and programmatic transformation of the data. In the logical mapping stage, a source-to-target (i.e., legacy to standard) specification was developed to provide rules for creating new, standardized data sets of the existing variables in the legacy data sets. This step could not be fully automated; it involved many person-hours and collaboration between data managers, programmers, and often even clinical subject matter experts to ensure that the clinical utility and meaning of the data were not compromised. These legacy data usually did not contain the related metadata or documentation to effectively understand the data without reviewing questions with a subject matter expert internal to the contributing organization to avoid confusion and error. Often, the process involved splitting or concatenating multiple variables, and separating variables that may have been grouped together in the legacy data into multiple data sets based on the target standard format (SDTM). In the programmatic transformation stage, clinical data programmers wrote scripts and programs to execute the plan described in the mapping specification to create the standardized data sets. In almost all cases, the data mapping was performed by the contributors.

Because of the potential for errors in the process, the final, crucial step in remapping was validation, to demonstrate the accuracy of the data. During this phase, data were checked for the conformance to the SDTM standard. CAMD data management used OpenCDISC™, an open source, freely available validation program to perform this task [12]. OpenCDISC checked for conformance to approximately 200 rules, including terminology, structure, and cross-dataset agreement. Additional checks were performed by CAMD for the conformance to new rules for terminology that had not yet been incorporated into the OpenCDISC software. The program generated validation reports, which CAMD then annotated with instructions or requests for clarification from the contributors. Validating and reconciling errors was often an iterative process working with the contributors to request changes and revalidating until data were suitable for the production database.

2.3. Data deidentification

Data in the CODR-AD database were deidentified in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor requirements [13]. Conforming to these guidelines entailed removing 18 so-called "identifiers" from the data, e.g., name, address, social security number, etc. It is important to note that most of these identifiers are not typically recorded in clinical trials data, and were therefore not supplied to CAMD. When available, and to ensure compliance with HIPAA Safe Harbor requirements, any age >89 years-of-age was converted to "999," whereas full year-month-day dates were first converted to an integer representing the number of days elapsed from each subject's reference start date (defined day 1).

2.4. Mixed effects modeling

The clinical trial simulation tool developed by CAMD using the CODR-AD database as one source of data is based on mixed effects population models (for disease progression, placebo effect and symptomatic drug effect), and a Weibull survival model for patient dropouts. As such, these approaches are suited to identify sources of variability that drive, for example, varying rates of disease progression within specific subpopulations, or the varying probability of patients dropping out. Only by integrating multiple data sources can such models help identify such subpopulations, and quantify the impact of such varying rates of progression on the design and analysis of clinical trials. This approach is supported by the results described by Rogers et al. [14], and in the endorsement decision from FDA [15] and EMA [16]. See Rogers et al. [14], for a detailed methodology of mixed effect model development.

This same concept applies to the placebo effects function, which captures relevant baseline sources of variability (baseline age and severity, gender and *APOE* ε4 genotype) that can help design teams envision scenarios for varying magnitudes, durations, and variability of placebo response. The placebo effects quantitative description allows researchers to envision scenarios regarding magnitude, duration, and variability of the placebo response, according to the selected entry criteria for the simulated trial [17]. As with any longitudinal modeling approach, variance always increases as a function of time. However, these quantitative tools are (by nature), continuously evolving entities that get continuously refined as additional data become available [17].

2.5. Consortium approach

A key factor critical to the success of the CODR-AD database was the use of a consortium approach to sharing data and information. All full-member organizations of CAMD assigned a representative to the CAMD Coordinating Committee. This committee determines the direction, budget, and policies of CAMD. Sharing clinical trials data, even when limited to the control arms, is not without perceived risk to the contributor. To mitigate this risk and address the concerns of the members, CAMD made the use of a consortium legal agreement and a separate data use agreement that spells out the acceptable use and access policies to the database. Additionally, the consortium legal agreement specifies that all publications produced by CAMD must be presented to the Coordinating Committee for review and input before being submitted for publication.

3. Results

To date, CAMD has received data on a total of 6500 subjects from 24 remapped studies of AD and mild cognitive impairment (MCI) from nine member organizations: Abbott (now AbbVie), the Alzheimer's Disease Cooperative Study, AstraZeneca, Eisai, Forest, GlaxoSmithKline, Johnson &Johnson, Pfizer, and Sanofi [17]. The diagnostic status of subjects in the database according to the stage of AD is MCI: n = 1041; moderate to severe AD: n = 146; severe AD: n = 377; mild to moderate AD: n = 4936.

A summary of the trials and baseline subject characteristics is shown in Table 2 and descriptive statistics of the subjects are represented in Fig. 1. Approximately 3200 subjects formed the analysis data set for an integrated approach by the CAMD modeling and simulation team to analyze the data as one key component for the development of a clinical trial simulation tool for mild and moderate AD [14]. The tool has been recently endorsed by both the EMA and FDA as the first drug-disease-trial model to achieve a regulatory decision [15,16]. In addition to CODR, the modeling tool incorporated patient-level data from ADNI and summary data from the literature. This quantitative drug development tool enables users to simulate phases 2 and 3 trials within the drug development process based on longitudinal ADAS Cog scores, and all their sources of variability, in mild and moderate AD patients (Fig. 2). Relevant covariates for disease progression include gender, number of *APOE* ε4 alleles, baseline age, and baseline disease severity (captured by the baseline MMSE score).

The CAMD consortium members agreed to make the CODR-AD database available to qualified external researchers. The rationale was to be sure to maximize the impact of the investments in the AD database beyond the primary goals of the consortium. At present, there are a growing number of examples of diverse research questions that are being addressed by analyzing the CODR-AD database (Table 3). The field is currently realizing the critical importance of data sharing to identify subtle signals in heterogeneous diseases; such strategies will serve to catalyze the concept of personalized medicine and de-risk drug development, an urgent need for AD.

4. Discussion

Although there are several other AD databases available to researchers, CODR-AD is unique in that it is the first

Table 2
Summary characteristics of trials and subjects in CODR-AD

STUDYID	Duration (weeks)	N	Female %	Years since DX	Background therapy
1000	12	102	58.8	2.5 (<1–13)	Yes
1009	12	164	55.5	0.9 (<1–11)	No
1013	78	719	50.2	2 (<1–10)	Both
1014	78	644	56.2	2.1 (<1–11)	Both
1055	52	140	58.6	NA*	No
1056	54	494	55.9	2.5 (<1–20)	Both
1057	54	500	61.4	2.1 (<1–10)	Both
1058	24	166	59	1.5 (<1–10)	No
1105	78	326	50.9	2.2 (<1–12)	Yes
1107	24	146	61	2.1 (<1–11)	No‡
1131	24	57	59.7	2.6 (<1–10)	No‡
1132	52	412	43.5	3.3 (<1–24)	No
1133	30	162	61.1	NA*	No‡
1134	24‡	105	81.9	NA*	No‡
1135	30‡	274	55.1	NA*	No‡
1136	52	144	59	NA*	No‡
1137	24	216	50.5	3.6† (<1–10)	Yes
1138	24	202	57.4	3.4† (<1–20)	No
1139	24	167	67.7	5.6 (<1–19)	No‡
1140	24	137	42.3	2.6 (<1–20)	No‡
1141	104	492	55.2	0.3 (<1–5)	No‡
1142	78	409	56	4.4 (<1–20)	Both
1143	24‡	105	82.9	5.4 (<1–20)	No‡
1144	54‡	217	64.5	3.6 (<1–13)	No‡
		N = 6500			

Abbreviation: CODR-AD, Coalition Against Major Diseases created an Online Data Repository for AD.

NOTE. *Study ID* is the unique identifier assigned to each study by CODR. N refers to the number of patients randomized to control arm contained in each study. *Years since DX* refers to the mean years because the diagnosis of AD or MCI at the start of each study, and ( ) contains the range in years. *Background therapy* identifies whether studies enrolled patients who were stably treated with either memantine, an aceteylcholinesterase inhibitor, or both at trial start; such therapy was neither an inclusion nor exclusion criterion for these studies, as it was in the case of studies marked "Yes" and "No", respectively.

*Data not available; could not be derived because the date of diagnosis was not provided.

†Studies 1137 and 1138: *Years since DX* was calculated based on a supplemental variable for estimated start of cognitive problem as collected in these studies, because a formal diagnosis date was not available in medical history.

‡These values were determined based on the presence or absence of acetylcholinesterase inhibitors or memantine in the data. Neither protocols nor clinicaltrials.gov listings were available to make this determination.

database available to qualified researchers that pools patient-level records from clinical trials data by adhering to an open-source standard. By pooling data in this fashion, analysts are able to query all trials or subsets of trials contained in the database without having to re-write programming statements for each new study. By providing longitudinal results from a variety of assessment tools, these studies enable researchers to better understand how the disease progresses and identify critical points along the disease continuum where intervention may be most effective.

Moreover, such data repositories align with the goals of C-Path and the FDA to use precompetitive data sharing as a means to improve efficiency in clinical trials. FDA created the Janus Clinical Trials Repository Project in 2010 to provide a hub for integrating data submitted to the agency as supporting evidence for regulatory decisions [18], and has begun converting legacy data and developing analytic tools to make these data more useful.

CODR, launched in 2010, currently supports several C-Path consortia, including CAMD, the Polycystic Kidney Disease Outcomes Consortium, and the Predictive Safety Testing Consortium, for a total of seven databases that contain data from nearly 150 studies (Fig. 3). C-Path's newest consortium, the Multiple Sclerosis Outcomes Assessment Consortium will also use CODR for sharing data from Multiple Sclerosis clinical trials [19]. Access to each database is managed separately according to the policies of its parent consortium. Among all C-Path CODR databases, the CODR-AD database is, so far, unique in being the only database that is available for external qualified researchers. Additional CODR databases may be made available to qualified external researchers in the future, dependent on the objectives of each C-Path consortium. The shared mission of C-Path consortia is to foster the development of drug development tools by precompetitive data sharing across member companies. The CODR database infrastructure represents a common means to accomplish this shared goal.

The CAMD CODR-AD database does not presently contain biomarker data (neuroimaging and biofluids such as cerebrospinal fluid analytes). The field of AD is struggling with the fact that there is a lack of consensus on the specific methodology and assay protocol standards employed in clinical trials to date using biomarkers. This is the case for both biofluid and neuroimaging biomarkers and is one of the many reasons why ADNI, which does use consensus protocols has been so successful. Although it is not possible at the present time to pool biomarkers across distinct randomized controlled clinical trials into a unified clinical trial database, the adoption of consensus Alzheimer's disease specific CDISC data standards in ongoing and future trials will positively impact the future.

At present, this database does contain clinical data representing mild to moderate stages of AD (~5500 subjects) and predementia trials at the stage of MCI (~1000 subjects). CODR-AD enables investigators not only to access and download data but also provides a web interface to analyze data with a commonly used open-source statistical program ("R") [20] and create and download reports. A web interface for generating reports via Structured Query Language is also provided.

Other data repositories have also been created for data from AD clinical trials: the Global Alzheimer's Association Interactive Network [21], a cloud-based, federated data repository of AD research data, which is currently under development by the Alzheimer's Association and the Laboratory for NeuroImaging at the University of Southern California [21]; a donepezil data repository that aggregates clinical trial

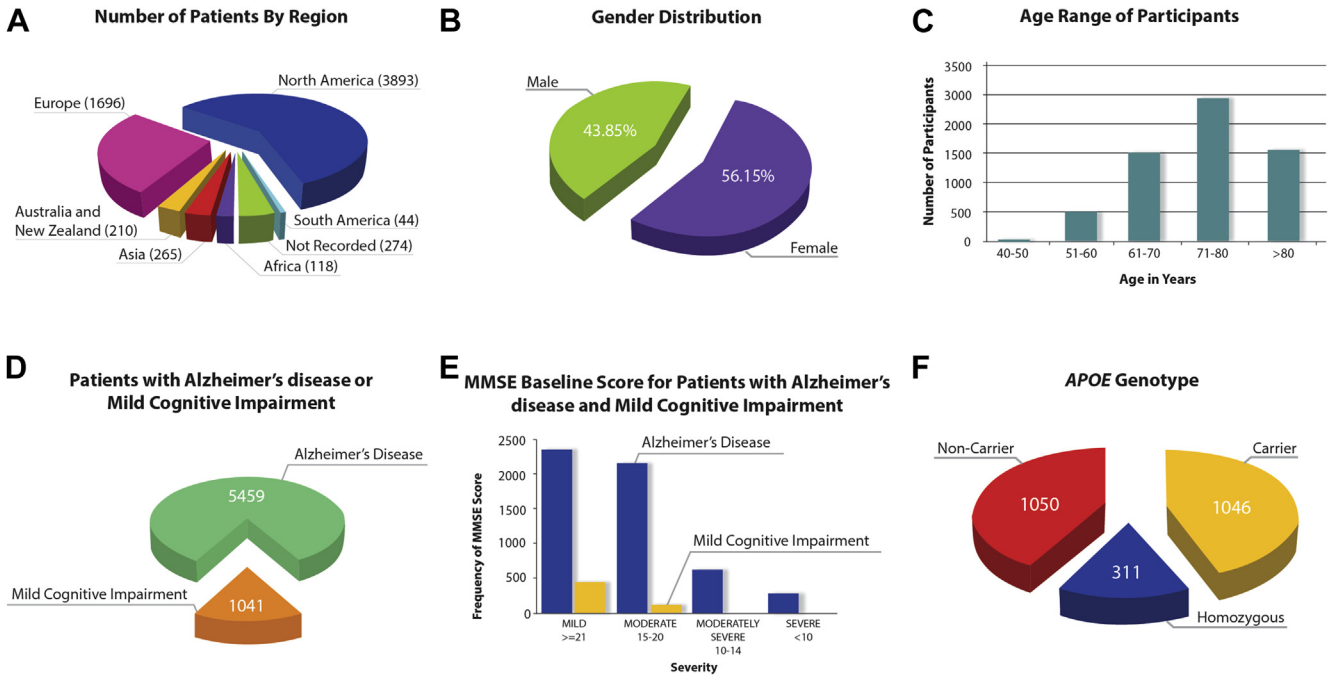


Fig. 1. Descriptive statistics for the subjects (N = 6500). Represented in the Coalition Against Major Diseases created an Online Data Repository for AD (CODR-AD) database. (A) Number of subjects by region. (B) Distribution of gender. (C) Categorical age range distribution. (D) Number of subjects by primary diagnosis at start of trial. (E) Number of subjects by baseline severity of cognitive deficit as measured by Mini-Mental Health Examination (MMSE) for both Alzheimer's disease (AD), blue bars) and mild cognitive impairment (MCI, yellow bars). (F) *APOE* genotypes represented in CODR-AD. Homozygous subjects are defined by having the ε4 isoform in both alleles. Carriers have one ε4 isoform allele. Noncarriers do not have the ε4 isoform in either allele.

data from 18 randomized, controlled trials conducted between 1991 and 2005 by Pfizer and Eisai [22], and the National Alzheimer's Coordinating Center database [23]. A number of other databases of longitudinal aging studies have also been created [24], yet these vary in terms of the types of data included and availability to other researchers.

It is important to note that data integration would not have been possible without the use of data standards. Overall, CAMD members contributed significant in-kind resources to remap all data including ADAS-Cog subscales to the AD CDISC standard. The successful development of the AD modeling tool could not have been achieved from a single sponsor. Pooling data sets as the CAMD team has done in the CODR-AD database created a unique and powerful research resource by enabling scientists to query for information across all data sets in the database simultaneously.

C-Path shares AD data with CAMD consortium members and qualified researchers who request access via the CODR website [24]. When requesting access, users are asked to send intended research questions and approaches to CAMD for approval. Once approved, users are able to query and analyze data relevant to their research questions. External disclosures are to be communicated to CAMD and publications acknowledge the consortium. To date, CAMD is aware of multiple external uses of the CODR database to address various research questions that have resulted in the publications outlined in Table 3 [17,25–30]. Additionally, we are aware of multiple abstracts that have been presented [31–33].

4.1. Caveats

Although the database has the potential to be very powerful, a number of caveats should be kept in mind. First, integrating data does not necessarily mean that those data are poolable from a statistical, scientific, or clinical standpoint. The database contains 24 studies, but it cannot be assumed that all 24 are suitable to answer every analysis question. Therefore, users must determine which studies are suitable for their analysis question(s). Furthermore, although data are standardized with regard to the SDTM variables and structure, some terminology was left verbatim as submitted by contributors. This is most notable in the labs data set where, for example, white blood cell counts may be referred to as "WBC" or "leukocytes." Also, the concomitant use of acetylcholinesterase inhibitors is a potentially major confounder to assessing outcomes, so CAMD asked members to adhere to generic drug names for this class of drugs when they were present in the data (as background therapy, for instance).

The primary cognitive outcome measure in AD is the ADAS-Cog, but this scale is not highly standardized, with 10-, 11-, 12-, and 13-item versions of the scale used by sponsors contributing data. Most trials include a common set of 11

web 4C/FPO

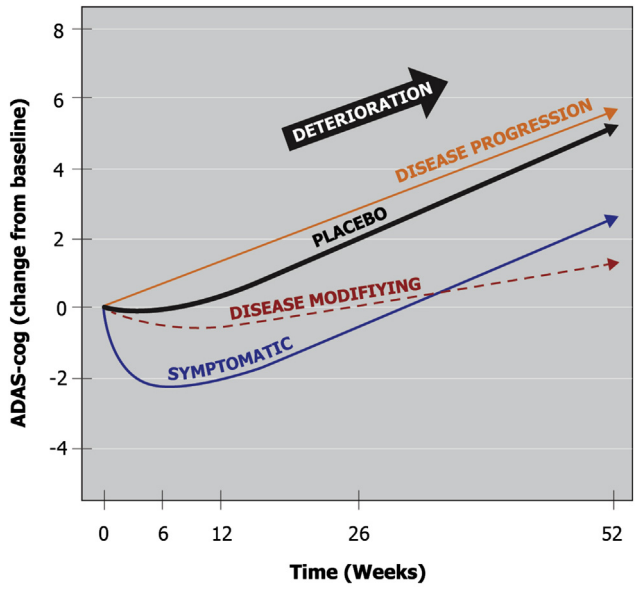


Fig. 2. Conceptual representation of the mild and moderate Alzheimer's disease (AD) CTS Tool. Standardized data from different sources provide the necessary information to develop quantitative models that capture the relevant aspects of disease progression, pharmacologic effects (which for AD have been categorized as "symptomatic" or "disease-modifying"), and aspects of trial design such as the magnitude duration and variability of the placebo effect. Such an integrated drug-disease-trial model forms the basis for the CAMD clinical trial simulation tool for mild and moderate AD. CTS, clinical trial simulation.

items from a total pool of 15 items. Moreover, the items do not always appear in the same order, and each sponsor may have had different rules in their respective analysis plans on how to handle missing data. Our resolution was to ask each member to provide the lowest level of raw detail on the individual items rather than their own analyses. If a given item was missing, they were instructed to populate the "status" variable as "not done" and then to define the "reason not done" as either due to "cognitive reasons" or "noncognitive reasons," if this information was available. This allows investigators the ability to dig deeper into each subject's performance and derive their own scores according to their own analysis plans. The supplementary materials include recommendations with regard to approaching such issues in analysis.

The authors acknowledge that there are numerous challenges that have plagued successful development of therapies in AD to date, only some of which may be addressed by sharing data from retrospective trials and use of predictive modeling and simulation tools. The field of AD is evolving in many ways such as new diagnostic guidelines that include biomarkers. The lack of biomarkers in the current CAMD-AD database does pose limitations in applicability to addressing some key research questions and it is acknowledged that the impact of the CAMD-AD database will be expanded with the inclusion of biomarkers.

5. Moving forward

The C-Path Online Data Repository for AD serves as an example of what can be achieved by standardization and integration of clinical trial database from industry-sponsored AD trials. The standardization and pooling of clinical trial data facilitates the analysis of data across multiple studies, providing a more comprehensive understanding of the disease process.

CODR-AD is an evolving database and the data standards developed as part of this project are not intended to be static. C-Path partnered with CDISC with the publication of v2.0 of the AD standards, which incorporates expanded clinical endpoints and biomarkers and is focused on predementia stages of AD (http://www.cdisc.org/therapeutic). CAMD is also working with partners to develop a more user-friendly access interface, and discussions are currently underway with other organizations, including ADNI, to develop approaches for the broader use of CDISC data standards, and integrated databases.

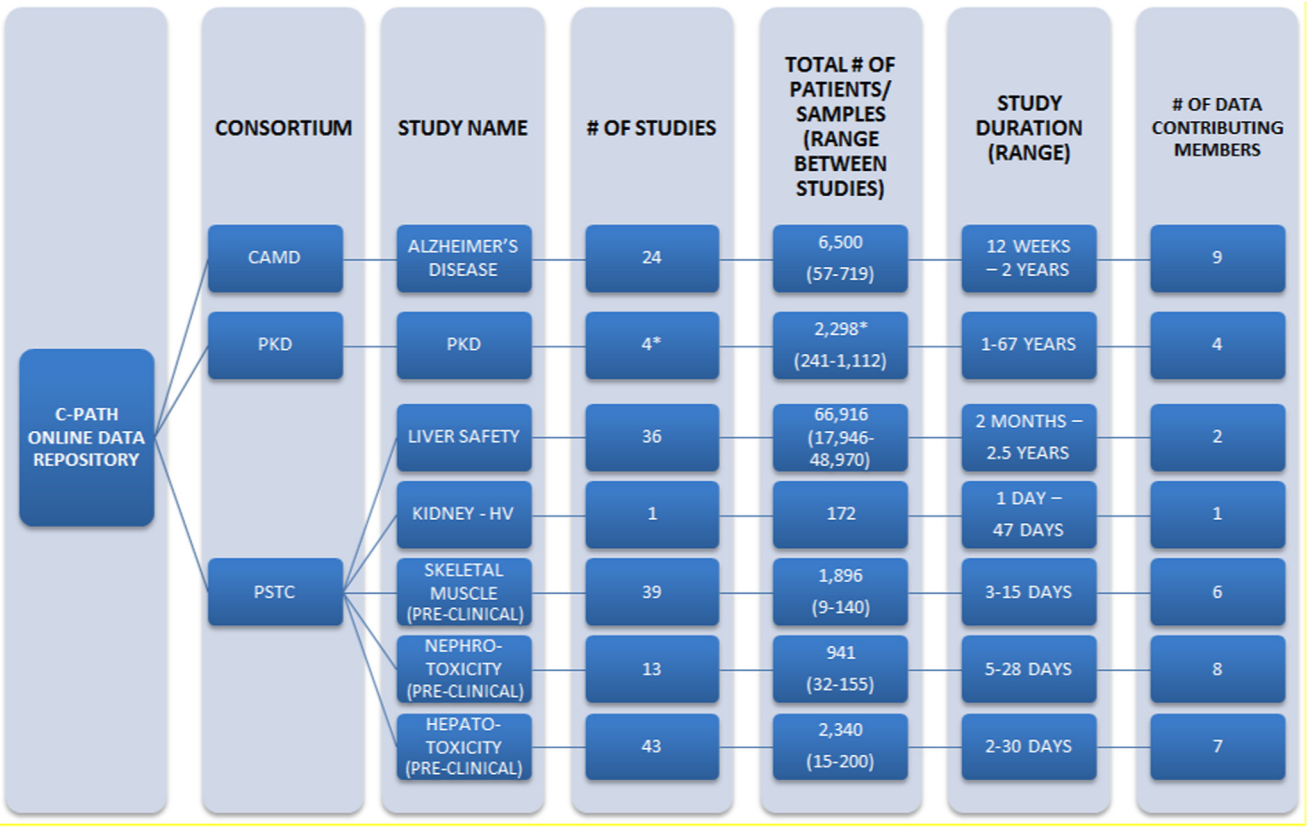
Table 3
Publications by external users of the CODR-AD database

Title	Research topic	Reference
Disease progression meta-analysis model in AD.	Disease progression modeling	Ito et al. [25].
Differences between early and late onset AD	Characterizing clinical features of early- vs. late-onset AD biomarkers research	Panegyres and Chen [29].
Identifying combinatorial biomarkers by association rule mining in the CAMD Alzheimer's database	Combinatorial biomarkers	Szalkai et al. [27].
Reliability of the ADAS-Cog in longitudinal studies	Interrater reliability; test-retest reliability; internal consistency	Khan et al. [28].
Understanding placebo responses in AD clinical trials from the literature meta-data and CAMD database.	Placebo response	Ito et al. [17].
Early-onset AD: a global cross-sectional analysis.	Characterizing early onset AD	Panegyres and Chen [26].
Improved utilization of ADAS-Cog assessment data through item response theory based pharmacometric modeling.	Item response theory	Ueckert et al. [30].

Abbreviations: CODR-AD, Coalition Against Major Diseases created an Online Data Repository for AD; AD, Alzheimer's disease; CAMD, Coalition Against Major Diseases; ADAS-Cog, Alzheimer's Disease Assessment Scale Cognitive subscale.

NOTE. A summary of the known publications based on the research utilization of the CODR-AD database by investigators external to CAMD. Abstracts also exist yet are not listed here.

web 4C/FPO



web 4C/FPO

Fig. 3. Contents of the C-Path online data repository. A breakdown of the studies contained in all Coalition Against Major Diseases created an Online Data Repository (CODR) databases across the participating C-Path consortia. The databases maintained by Coalition Against Major Diseases (CAMD) and the Polycystic Kidney Disease Outcomes Assessment Consortium (PKD) make use of Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM). Predictive Safety Testing Consortium (PSTC) studies are primarily nonclinical. The contents of the database are accurate as of the publication of this article. The number of new studies and working groups among consortia are dynamic and subject to growth. HV, healthy volunteer study.

Other data will continue to be incorporated into CODR-AD as it becomes available. CAMD is working with companies to develop a framework under which they would be willing to make more data available, such as inclusion and exclusion criteria, trial design methods, active treatment arm data, and biomarkers across the disease continuum. Although companies have expressed concerns that by making this information available, (i.e., they could risk losing their anonymity as contributors), regulators and many sponsors agree that these data are essential for more efficient analysis and interpretation of the database [34]. CODR-AD database and its use to date serves as an example that responsible use and effective and impactful advances will emerge from big data. The landscape of precompetitive data sharing is changing in a positive way and the CODR-AD database serves as an example for others interested in big data across disciplines and diseases.

Acknowledgments

The authors thank Dr. Arthur Toga (Laboratory of Neuro Imaging, Keck School of Medicine of University of Southern California) for review of the manuscript. We also gratefully acknowledge the contributions of Dr. Ray Woosley, Rick Myers, Andy Tofel and Aaron Avery (Ephibian), Dr. Rebecca Kush (CDISC), Cathy Barrows (GSK), Roberta Rosenburg and Ann Pennington and Kaori Ito (Pfizer), Suzanne Pierre (Sanofi), and Chris Tolk (CDISC).

The success of this database and goals of the CAMD consortium would not have been possible without the contribution of subject-level clinical trials data. We wish to thank Sanofi, Pfizer, GlaxoSmithKline, Johnson &Johnson, Abb-Vie, Forest Laboratories, Eisai, AstraZeneca, and the Alzheimer's Disease Cooperative Study (ADCS) for their data contributions.

Funding: Partial funding for this work came from grants from Science Foundation Arizona (SRG 0335-08) and the U.S. Food and Drug Administration (5U01 FD003865).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2014.11.005.

RESEARCH IN CONTEXT

1. Systemic review: The data team from the Coalition Against Major Diseases surveyed available clinical databases for Alzheimer's disease (AD) research and assessed the approaches used by the developers of those databases. Additionally, the team surveyed available clinical data standards (in particular, standards specific to AD) suitable for creating integrated databases.
2. Interpretation: A standardized database of clinical trials on AD (predementia and dementia)—the first of its kind—was developed and made available to qualified researchers. A new open-source, publicly available CDISC data standard for Alzheimer's predementia and dementia was also developed and is available for use in prospective clinical studies.
3. Future directions: Maximizing the usefulness of the database will require incorporating data from additional trials, including biomarker data and data from other outcome assessments and endpoints. The authors are actively seeking these data and will be updating data standards that will enable their incorporation into the database.

References

[1] Alzheimer's Disease International. World Alzheimer's report. 2009.
[2] Olson S, Downey AS. Sharing clinical research data: workshop summary. Washington, D.C: Institute of Medicine of the National Academies. 2013.
[3] Food and Drug Administration. Innovation or stagnation: challenges and opportunity on the critical path to new medical products. 2004. Available at: http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/CriticalPathOpportunitiesReports/ucm077262.htm. Accessed February 26, 2013.
[4] Romero K, Corrigan B, Neville J, Kopko S, Cantillon M. Striving for an integrated drug development process for neurodegeneration: the coalition against major diseases. Neurodegen Dis Manage 2011;1:1–7.
[5] Romero K, de Mars M, Frank D, Anthony M, Neville J, Kirby L, et al. The coalition against major diseases: developing tools for an integrated drug development process for Alzheimer's and Parkinson's diseases. Clin Pharmacol Ther 2009;86:365–7.
[6] Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. Alzheimers Dement 2013;9:e111–94.
[7] Clinical Data Interchange Standards Consortium. Clinical Data Interchange Standards Consortium.
[8] Food and Drug Administration. PDUFA reauthorization performance goals and procedures fiscal years 2013 through 2017. 2012. Available at: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM270412.pdf. Accessed June 27, 2013.
[9] Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–3.
[10] Gobburu JV, Lesko LJ. Quantitative disease, drug, and trial models. Annu Rev Pharmacol Toxicol 2009;49:291–301.
[11] Clinical Data Interchange Standards Consortium. Alzheimer's disease-specific therapeutic area supplement to the study data tabulation model user guide. Prepared by the Coalition Against Major Diseases. Jan 2014. Available at: http://www.cdisc.org/system/files/all/standard_category/application/pdf/taug_alzheimer_27s_20v2.0_20noportfolio.pdf. Accessed February 11, 2015.
[12] OpenCDISC. OpenCDISC Validator. Available at: http://www.opencdisc.org/. Accessed February 11, 2015.
[13] U.S. Department of Health and Human Services. Guidance regarding methods of de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. Available at: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html. Accessed February 11, 2015.
[14] Rogers JA, Polhamus D, Gillespie WR, Ito K, Romero K, Qiu R, et al. Combining patient-level and summary-level data for Alzheimer's disease modeling and simulation: a beta regression meta-analysis. J Pharmacokinet Pharmacodyn 2012;39:479–98.
[15] Food and Drug Administration. Regulatory letter to critical path institute and CAMD regarding Alzheimer's disease simulation tool. FDA Center for Drug Evaluation and Research, Disease specific model library. Available at: http://www.fda.gov/aboutfda/centersoffices/officeofmedicalproductsandtobacco/cder/ucm180485.htm. Accessed February 11, 2015.
[16] European Medicines Agency. Qualification opinion of a novel data driven model of disease progression and trial evaluation in mild and moderate Alzheimer's disease. 2013. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2013/10/WC500151309.pdf. Accessed February 11, 2015.
[17] Ito K, Corrigan B, Romero K, Anziano R, Neville J, Stephenson D, et al. Understanding placebo responses in Alzheimer's disease clinical trials from the literature meta-data and CAMD database. J Alzheimers Dis 2013;37:173–83.
[18] Food and Drug Administration. Janus Clinical Trials Repository (CTR) Project 2012 [6/27/13]. Available at: http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm155327.htm. Accessed February 11, 2015.
[19] Rudick RA, Larocca N, Hudson LD. Multiple Sclerosis Outcome Assessments Consortium: genesis and initial project plan. Mult Scler 2013;20:12–7.
[20] R Project for Statistical Computing. Available at: http://www.r-project.org/. Accessed February 11, 2015.
[21] Alzheimer's Association. Global Alzheimer's Association Interactive Network 2013. Available at: http://www.gaain.org/. Accessed February 11, 2015.
[22] Jones R, Wilkinson D, Lopez OL, Cummings J, Waldemar G, Zhang R, et al. Collaborative research between academia and industry using a large clinical trial database: a case study in Alzheimer's disease. Trials 2011;12:233.
[23] Beekly DL, Ramos EM, Lee WW, Deitrich WD, Jacka ME, Wu J, et al. The National Alzheimer's Coordinating Center (NACC) database: the uniform data set. Alzheimer Dis Assoc Disord 2007;21:249–58.
[24] CAMD. C-Path online data repository. Available at: https://codr.c-path.org. Accessed February 11, 2015.
[25] Ito K, Ahadieh S, Corrigan B, French J, Fullerton T, Tensfeldt T, et al. Disease progression meta-analysis model in Alzheimer's disease. Alzheimers Dement 2010;6:39–53.
[26] Panegyres PK, Chen HY, the Coalition against Major Diseases. Early-onset Alzheimer's disease: a global cross-sectional analysis. Eur J Neurol 2014 Apr 30; http://dx.doi.org/10.1111/ene.12453 [Epub ahead of print]. 2014.
[27] Szalkai B, Grolmusz VK, Grolmusz VI, Coalition Against Major Diseases. Identifying combinatorial biomarkers by association rule mining in the CAMD Alzheimer's database. Cornell University Library.

ARTICLE IN PRESS

2013. Available at: http://arxiv.org/abs/1312.1876. Accessed February 11, 2015.

[28] Khan A, Yavorsky C, DiClemente G, Opler M, Liechti S, Rothman B, et al. Reliability of the Alzheimer's disease assessment scale (ADAS-Cog) in longitudinal studies. Curr Alzheimer Res 2013;10:952–63.

[29] Panegyres PK, Chen HY. Differences between early and late onset Alzheimer's disease. Am J Neurodegener Dis 2013;2:300–6.

[30] Ueckert S, Plan EL, Ito K, Karlsson I, Corrigan B, Hooker AC. Improved utilization of ADAS-Cog assessment data through item response theory based pharmacometric modeling. Pharm Res 2014; 31:2152–65 [Epub ahead of print].

[31] Ueckert S, Plan EL, Ito K, Karlsson MO, Corrigan B, Hooker AC. Application of item response theory to ADAS-cog scores modelling in Alzheimer's disease. Venice, Italy: Population Approach Group in Europe; 2012. p. 21.

[32] Yavorsky C, DiClemente G, Opler M, Khan A, Jovic S, Rothman B. Establishing threshold scores and profiles of cognitive impairment for the Alzheimer's Disease Assessment Scale— Cognitive Subscale (ADAS-Cog) for patients with higher dementia (MMSE < 12), Alzheimer's disease and probably MCI. Alzheimers Dement 2012;8:P415–6.

[33] Khan A, Yavorsky C, Opler M, Jovic S. Differential item functioning and the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) among people with Alzheimer's disease. Alzheimer Dement 2013;9:P460–1.

[34] Eichler HG, Petavy F, Pignatti F, Rasi G. Access to patient-level trial data–a boon to drug developers. N Engl J Med 2013;369:1577–9.