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.

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

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