frontiers in Molecular Biosciences

REVIEW
published: 14 January 2020
doi: 10.3389/fmolb.2019.00158

# Challenges of Integrative Disease Modeling in Alzheimer's Disease

*Sepehr Golriz Khatami*[1,2]*, *Christine Robinson*[1,2], *Colin Birkenbihl*[1,2], *Daniel Domingo-Fernández*[1,2], *Charles Tapley Hoyt*[1,2] and *Martin Hofmann-Apitius*[1,2]

[1] Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, Sankt Augustin, Germany,
[2] Bonn-Aachen International Center for IT, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

Dementia-related diseases like Alzheimer's Disease (AD) have a tremendous social and economic cost. A deeper understanding of its underlying pathophysiologies may provide an opportunity for earlier detection and therapeutic intervention. Previous approaches for characterizing AD were targeted at single aspects of the disease. Yet, due to the complex nature of AD, the success of these approaches was limited. However, in recent years, advancements in integrative disease modeling, built on a wide range of AD biomarkers, have taken a global view on the disease, facilitating more comprehensive analysis and interpretation. Integrative AD models can be sorted in two primary types, namely hypothetical models and data-driven models. The latter group split into two subgroups: (i) Models that use traditional statistical methods such as linear models, (ii) Models that take advantage of more advanced artificial intelligence approaches such as machine learning. While many integrative AD models have been published over the last decade, their impact on clinical practice is limited. There exist major challenges in the course of integrative AD modeling, namely data missingness and censoring, imprecise human-involved priori knowledge, model reproducibility, dataset interoperability, dataset integration, and model interpretability. In this review, we highlight recent advancements and future possibilities of integrative modeling in the field of AD research, showcase and discuss the limitations and challenges involved, and finally, propose avenues to address several of these challenges.

**Keywords: Alzheimer's disease, challenges, integrative disease modeling, hypothetical, data-driven**

OPEN ACCESS

***Edited by:***
*Daniel Garrido, Pontifical Catholic University of Chile, Chile*

***Reviewed by:***
*Lin Tan, Qingdao Municipal Hospital, China*
*Owen T. Carmichael, Pennington Biomedical Research Center, United States*

***Correspondence:**
*Sepehr Golriz Khatami*
*sepehr.golriz.khatami@scai-extern.fraunhofer.de*

***Specialty section:***
*This article was submitted to Biological Modeling and Simulation, a section of the journal Frontiers in Molecular Biosciences*

***Received:*** *02 September 2019*
***Accepted:*** *18 December 2019*
***Published:*** *14 January 2020*

***Citation:***
*Golriz Khatami S, Robinson C, Birkenbihl C, Domingo-Fernández D, Hoyt CT and Hofmann-Apitius M (2020) Challenges of Integrative Disease Modeling in Alzheimer's Disease. Front. Mol. Biosci. 6:158. doi: 10.3389/fmolb.2019.00158*

## INTRODUCTION

Alzheimer's Disease (AD) manifests in a collection of symptoms including the deterioration of cognition, memory, and behavior which often leads to interference with activities of daily living. In 2017, AD ranked among the top five causes of death worldwide, with 2.44 million (4.5%) deaths from AD[1,2]. Worldwide, there are currently around 50 million people living with AD, and every 3 s a person develops this condition. It is estimated that only a quarter of those living with AD are diagnosed, and more than 17 million healthcare workers annually invest 18 billion hours of care, at a cost of more than one trillion US dollars to tackle AD-associated problems[3,4]. Extrapolating these statistics to the coming decades suggests the immense socioeconomic impact of AD on all involved

[1] https://ourworldindata.org/causes-of-death
[2] https://www.thestreet.com/world/leading-causes-of-death-world-14869811
[3] https://www.alz.co.uk/research/statistics
[4] https://ourworldindata.org/causes-of-death

parties: patients, caregivers, healthcare systems, and indirectly, the economy. Thus, strategies to reduce the global emotional and financial burden of AD are of great importance. To develop such strategies, a deeper understanding of the pathophysiology underlying AD is necessary and may lead to opportunities for earlier detection and therapeutic interventions.

In general, AD progression is categorized into three clinical disease stages: (i) During the pre-symptomatic phase, individuals may have already developed pathological changes that underlie AD, but remain cognitively normal, (ii) in the prodromal phase, often referred to as mild cognitive impairment (MCI), the first cognitive symptoms, commonly episodic memory deficits, appear. These symptoms can be acute, but they do not yet meet the criteria for dementia, (iii) in the dementia stage, impairments are severe enough to interfere with daily life (Jack et al., 2010).

Understanding of the etiology of AD is complicated due to the existence of dysregulations at different biological scales, ranging from genetic mutations to structural and functional alterations of the brain (Aisen et al., 2017). For this reason, significant efforts have been made in recent years to discover candidate markers for disease-related pathological changes throughout all modalities, including neuro-imaging, cerebrospinal fluid (CSF) samples and a broad variety of -omics data. Studies have successfully identified multiple biomarkers for neurodegeneration and AD (Blennow and Zetterberg, 2018). However, effectively translating extensive biomarker screenings into clinical application remains a challenging task, because individual biomarkers can only provide a highly incomplete view on such a multifactorial disease (Younesi and Hofmann-Apitius, 2013). For instance, while multiple associations between genetic variants and AD have been established (Jansen et al., 2019; Kunkle et al., 2019), none of these associations fully describe disease pathogenesis. As a result, one of the major challenges in AD research is translating diverse biomarker signals available into multimodal, multiscale models of disease pathogenesis.

In recent years, a new translational research paradigm called "integrative disease modeling" has emerged, to address this challenge (Younesi and Hofmann-Apitius, 2013). It aims at modeling heterogeneous measurements across different biological scales, in order to provide a holistic picture of biomarker intercorrelations in the disease of study. To this end, advanced high-throughput technologies and neuroimaging procedures are being used to collect data from multiple modalities. These diverse data need to be integrated, that is, combined in a way that preserves the structure and meaning in the data, using computational algorithms. Only then can they provide a solid basis for further analysis such as reasoning, simulation, and visualization. In order to contribute to understanding of the complex pathophysiology of the disease, the results should be actionable and thus must be interpretable. Integrative disease modeling, by collecting, integrating, analyzing, and ultimately interpreting the measurements, facilitates the understanding of the pathophysiology of complex diseases like AD (Hampel et al., 2017).

Existing integrative models in the context of AD can be placed in two primary categories, namely hypothetical models and data-driven models (**Table 1**). Hypothetical models are non-numerical and rely on reasoning over findings of previously published studies (Jack et al., 2010), rather than large amounts of data. By including this prior knowledge, these models try to detail the temporal changes of AD biomarkers relative to each other as well as to clinical disease stages and trial endpoints.

**TABLE 1 |** Organization of and references for data-driven integrative AD models.

| Data-driven integrative AD models | | | References |
|---|---|---|---|
| Traditional | | | Caroli and Frisoni, 2010; Jack et al., 2011, 2012 |
| Machine learning | Generative | | Fonteijn et al., 2012; Chen et al., 2016; Khanna et al., 2018; Oxtoby et al., 2018; Basu et al., 2019; De Jong et al., 2019; Gootjes-Dreesbach et al., 2019; Martinez-Murcia et al., 2019 |
| | Discriminative | Supervised | Hinrichs et al., 2010; Magnin et al., 2010; Rao et al., 2011; Zhang et al., 2011; Da et al., 2013; Li et al., 2013 |
| | | Unsupervised | Nettiksimmons et al., 2014; Gamberger et al., 2017; Toschi et al., 2019 |

*We subdivide data-driven integrative AD models which into two subgroups. While the first group uses simple statistical approaches (e.g., simple linear models), the second group uses more advanced techniques (e.g., machine learning). The advanced machine learning models include generative and discriminative models, the latter of which can be classified as either supervised or unsupervised models.*

By contrast, data-driven integrative models take advantage of developments in computational approaches and big data. For the sake of this review, we will distinguish between two subcategories of data-driven models. The first covers traditional statistical methods of generally lower complexity, such as linear models. Often, these models are used to estimate biomarker trajectories by regressing measured data against a prespecified dependent variable, such as a clinical readout or the disease stage (Bateman et al., 2012). The second subtype exploits more advanced artificial intelligence approaches such as machine learning. Within this subtype, models can be characterized as discriminative or generative. Discriminative models are designed to discriminate between groups (e.g., cases and controls) and can be further described as supervised or unsupervised, depending on whether they rely on labeled (Hinrichs et al., 2011; Da et al., 2013) or unlabeled (Toschi et al., 2019) data. Generative models contribute to disease understanding by automatically learning the inherent distribution of a dataset and its feature interdependencies (Oxtoby et al., 2018). An exemplary application is the extraction of disease progression signatures as demonstrated by the ensemble of Bayesian networks developed by Khanna et al. (2018).

Integrative AD modeling faces many challenges. Hypothetical models, by their nature, are time-intensive to construct and require specialist knowledge. Their primary role in AD research is to provide ideas for future experiments. Likewise in data-driven modeling, several challenges at each step of the process (i.e., collection, integration, analysis, and interpretation) must be addressed. Data missingness and data censoring are significant bottlenecks in data collection as well as analysis and

interpretation. Meanwhile, the heterogeneity and complexity of biological data are major impediments to data integration, which forms the basis for all data-driven approaches. Furthermore, data mapping, data labels, and biased data are additional barriers to robust data analysis and interpretation. Finally, insufficient numbers of subjects restrict the statistical power of data-driven integrative AD models. These fundamental challenges explain why, at this point in time, although many integrative AD models have been published over the last decade, their impact on clinical practice is limited.

In this review, we highlight recent advancements and future possibilities of integrative modeling, discuss the limitations and challenges involved, and finally, propose avenues to address several of these challenges, in the context of AD research.

## INTEGRATIVE AD MODELS

As already mentioned, integrative AD models can be characterized as either hypothetical or data-driven, each of which has strengths and weaknesses. In the following, we compare different models of each type and discuss their benefits and limitations. Finally, we elaborate on how associated limitations and challenges could be handled.

### Hypothetical Models

In hypothetical modeling, a model is generated about an object of study, direct knowledge of which is difficult to obtain. These models provide hypotheses about the object (Gladun, 1997). In integrative AD modeling, researchers develop so-called cascade models, in which the measurements of a set of biomarkers are normalized and their trajectories are plotted on a common time scale, aligned to disease stages (Jack et al., 2010, 2013). These models are typically developed by reviewing the available knowledge and reasoning over observations from previously published studies. They are not directly informed by measured data.

One of the first hypothetical integrative AD models was developed by Jack et al. (2013) [revised from a previous model (Jack et al., 2010)]. This model hypothesized the temporal changes of the five most studied biomarkers of AD pathology in relation to estimated years from expected symptom onset and in relation to other biomarkers. These biomarkers are CSF amyloid-beta protein (CSF $A\beta_{1-42}$) and tau protein (CSF tau) levels, amyloid-beta PET imaging (PET Aβ), Fluorodeoxyglucose-PET imaging, and structural MRI readouts. In this cascade model, the authors presumed that biomarker trajectories should exhibit a sigmoid-shaped curve. This imposition is a direct result of the limited sensitivity of measurements at time extremes, which the authors addressed by taking the floor of the measurements at early timepoints, and the ceiling of the measurements at late timepoints. The authors hypothesized that the two amyloid-beta (Aβ) biomarkers (i.e., CSF $A\beta_{1-42}$ and PET Aβ imaging) gradually approach an abnormal state while the subject remains in a cognitively normal state. After a lag period, the length of which varies from patient to patient, and in later disease stages, CSF tau, Fluorodeoxyglucose-PET, and structural MRI biomarkers follow the same pattern and begin the transition to an abnormal state. Similarly, Frisoni et al. (2010) established a theoretical progression of cognitive and biological markers (primarily imaging features) based not only on the clinical disease stages, but also patient age at AD diagnosis and time since diagnosis. Although both models captured earliest detectable changes in amyloid markers, Frisoni et al. (2010) additionally theorized that these changes plateau by the MCI stage, when the individuals are no longer cognitively normal. Furthermore, they suggested that F-fluorodeoxyglucose PET is abnormal by the MCI stage and continues to change well into the dementia stage. Structural changes appear later, following a temporal pattern mirroring tau pathology deposition, which slightly differs from the Jack et al. models (Jack et al., 2010, 2013).

While hypothetical models cannot be directly applied, they can be used to suggest directions for future experiments that themselves would address diagnosis, prediction, or decision making tasks (Gladun, 1997). However, there are a number of challenges relating to the construction of hypothetical models. In the following, we discuss these challenges and propose ways to address some of them.

### Challenges of Hypothetical Models

The exclusive reliance of hypothetical models on literature presents several challenges. First, relevant literature must be identified. Second, the scientific knowledge contained in the literature must be extracted in a meaningful form. Finally, the knowledge has to be modeled.

In order to build a hypothetical model, a researcher must identify a set of relevant publications, called a literature corpus, which accurately reflects AD knowledge. This corpus should be representative of the relevant aspects of AD, contain the most up-to-date publications, and not be biased toward subfields or trends. However, the number of new AD publications has increased each year since 2005, and there were nearly 15,000 such publications in 2017 alone (Dong et al., 2019). With such publication rates, it is challenging for researchers to manually create high quality corpora (Rodriguez-Esteban, 2015), Moreover, manual generation of these corpora is susceptible to bias, because researchers may tend to draw more heavily from authors or subfields with which they are more familiar (Atkins et al., 1992). The size of a corpus will also be limited by the time and resources available to the researchers. However, text mining has been used effectively to automatically classify relevant literature, based on titles and abstracts (e.g., see Simon et al., 2018), and to prioritize texts (Singh et al., 2015). Publications identified by this classification can be directly taken as the corpus or used as a more manageable set of publications from which the domain experts can appropriately select. Hypothetical models are susceptible to biases present in the literature (Boutron and Ravaud, 2018), but a well-designed, computationally selected corpus can mitigate the effects of those biases.

Once the corpus has been identified, the challenge of knowledge extraction remains. The goal here is to recover the knowledge contained in the publications in a meaningful way. Conducting this task manually is a time-consuming process that requires a high degree of domain knowledge. Here, text mining

poses the opportunity to extract knowledge in a computable form (Gyori et al., 2017; Lamurias and Couto, 2019). Moreover, it significantly reduces the amount of time required to read publications, which enables significantly larger corpora to be used in the building of hypothetical models.

Finally, in order to build hypothetical models, the information gleaned from the literature corpus must be organized in a coherent way. The entities and the relationships between them should all be represented. Mind maps provide a non-automated way of generating a knowledge model, driven by domain-expert knowledge (Kudelic et al., 2011). However, if automated information extraction strategies were used on the literature corpus, then knowledge graphs are well-suited for storing the extracted knowledge (Gyori et al., 2017). A major advantage of this strategy is that the knowledge graph is computable, meaning downstream machine learning tasks can be carried out for knowledge discovery. Furthermore, knowledge graphs support hypothesis generation by enabling researchers to assess whether their hypotheses are compatible with existing knowledge (Humayun et al., 2019).

Automated methods of corpus identification, knowledge extraction, and knowledge modeling provide a means of mitigating the challenges of hypothetical modeling. They reduce the time burden, mitigate the risk of bias in manual methods, and generate computable knowledge representations. This can yield more reliable hypothetical AD models.

Hypothetical models are non-numerical and rely exclusively on qualitative information, gleaned from a review of previous findings. This limits their usability solely to eliciting hypotheses for future experiments. They are neither predictive nor can they be used for analysis of any kind of data. They are meant to represent a kind of "typical" AD progression, without reflecting individual deviations from that. Given the broad biological heterogeneity observed among AD subjects, and the increasing relevance of personalized medicine (Reitz, 2016), there is a need for models that are capable of achieving this.

Data-driven models built on data collected in longitudinal cohort studies can serve to support or challenge hypotheses generated by hypothetical models (Petrella et al., 2019). Data-driven models are appropriate for a wide range of tasks that lie beyond the scope of what hypothetical models are designed for. For example, using data models can capture individual subject particularities that hypothetical models cannot (see e.g., Young et al., 2015). In the following, we discuss data-driven models and their challenges in depth.

## Data-Driven Models

In contrast to hypothetical models, data-driven integrative models are directly derived from datasets comprising readouts of multiple biomarkers. Such models can be applied to a broad variety of tasks ranging from predictive modeling e.g., predicting patient diagnosis (Ding et al., 2018) or age at disease onset (Chuang et al., 2016; Peng et al., 2016) to discovering patterns in the data that shed light on biomarker interdependencies and disease underlying mechanisms. Since these models use extensive data, they are not limited by preconceived notions in the way that hypothetical integrative models are.

Data-driven AD models can be classified into two primary subtypes based on the statistical approaches and algorithms applied (**Table 1**). The first subtype use traditional statistical methods such as linear modeling, and the second employs artificial intelligence and more specifically machine learning approaches.

### Traditional Statistical Models

In AD modeling, traditional statistical approaches, such as linear mixed-effects models, are often used to estimate biomarker trajectories (Caroli and Frisoni, 2010; Jack et al., 2011, 2012). In these models, measured data, are regressed against a prespecified variable, such as disease stage, to detail the temporal changes of AD biomarkers during the course of disease. Essentially, these models provide empirical testing of hypothetical multiple biomarker trajectory plots.

Jack et al. (2012) used linear mixed-effects models to investigate the shape of five important AD biomarker trajectories (i.e., $A\beta_{42}$, tau, amyloid, fluorodeoxyglucose PET, and structural MRI) as a function of a cognitive test score, the Mini-Mental State Exam (MMSE). This model parameterization enabled them to assess within-subject rates of biomarker changes with respect to changes of the MMSE score. They found that lower baseline MMSE scores are correlated with worse baseline biomarker values and that higher rates of biomarker change were associated with worsening MMSE score. This model constructed the biomarker trajectories without making any assumptions about the shapes of the trajectories. This contrasts with the authors' earlier hypothetical biomarker cascade model, which imposed a sigmoid trajectory curve.

While the shapes of the trajectories in this data-driven model agree with the assumptions made in the hypothetical exemplar, the model has several limitations, pertaining to model design choices and deficiencies in the data. The authors chose to use the MMSE score as the independent variable. This choice was made because the MMSE score provides a linear measure of disease progression that was available across all datasets. However, this introduces challenges in the estimation of trajectories in early disease stages, because MMSE scores in cognitively normal patients are relatively stable over time (Tombaugh, 2005), yielding only a narrow range of values. Moreover, especially when studying early disease stages, the model additionally suffers from possible absence of information on future disease developments of a subject. This absence of data on future disease outcome is related to data censoring, which will be addressed in more detail later.

In their data-driven model (Jack et al., 2011), Jack et al. aimed to unravel the temporal order of biomarker trajectories becoming abnormal, rather than only describing the shape of their trajectories. They used the prevalence of biomarker abnormalities at different disease stages to empirically assess the temporal ordering of their trajectories. They employed generalized estimating equations, a generalized linear model for longitudinal data that can deal with correlated observations, to evaluate and compare the proportion of abnormal observations per biomarker. The proper choice of a cut-off defining when biomarker measures are considered to be abnormal is a point

of debate and making this choice requires critical judgement. To differentiate between normal and abnormal biomarkers, Jack et al. (2011) determined a cut-off by looking at an independent post-mortem cohort. However, since, by construction, results were highly sensitive to the selected cut-off for each biomarker, the temporal resolution of the model is limited.

While the proportion of patients with abnormal biomarker values might seem an unnatural choice for comparing biomarkers, alternative strategies also have drawbacks. Caroli and Frisoni (2010) computed Z-scores based on values of each biomarker and fitted them against Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-cog) scores, comparing linear and sigmoidal fits. Their investigation showed that a sigmoid curve fit the observed data significantly better than a linear one for most of the biomarkers, and thereby might be able to characterize the time course of those biomarkers. These results were consistent with the hypothetical model proposed by Jack et al. (2010) and Jack et al. (2013). However, the biomarker trajectories cannot be directly compared with the data-driven model developed by Jack et al. (2011), since different scales were employed in both studies. While standardization of values by converting them into Z-scores resolves this problem, it introduces a new one: by definition, the arithmetic mean of each biomarker will be 0. This makes it impossible to reasonably compare biomarker distributions based on their means using standard statistical procedures like, for example, *t*-tests (Jack et al., 2011; Moeller, 2015).

The arbitrariness of defining a cut-off for abnormality of a biomarker will always pose a limitation on statistical approaches relying on biomarkers. While such cut-offs simplify the interpretation of the biomarker, there is no universally correct cut-off for a given biomarker. Rather, appropriate cut-offs heavily depend on the population, and even the individual, on which a biomarker will be used. Covariates such as an individual's age, genetic risk factors, and family history of AD must be considered. For these reasons, there is no single optimal cut-off for any given biomarker (Bartlett et al., 2012; Anne and Fagan, 2014). To address this, a less rigid technique has been developed, that designates an intermediate range using two cut-offs, one permissive and the other conservative (Klunk et al., 2012; Jack et al., 2016a,b; Bzdok, 2017). The permissive point can be used for earliest detectable evidence of AD pathologic changes and the conservative one for high diagnostic certainty. Moreover, different statistical approaches, like Youden's index and the receiver operating characteristic (ROC) curve, can be applied to help determine an appropriate cut-off.

Linear traditional models are ill-equipped to handle the increasingly high-dimensional data being collected in AD studies. Thanks to recent technological advancements, the granularity of AD datasets with respect to information resolution, feature size, and complexity of meta-information have increased. For example, improved neuro-imaging techniques generate datasets with higher resolution than previously available. This information distributed over voxels, a 3D imaging unit, is hard to capture using linear models (Bzdok, 2017). Therefore, more advanced data-driven models have been developed based on machine learning. These models are generally more flexible and compatible with the complex datasets encountered in biology research (Bzdok, 2017).

## Machine Learning Models

Machine learning models can be characterized as generative or discriminative. As previously mentioned, discriminative models are designed to differentiate between groups, while generative models provide better disease understanding by learning inherent properties from datasets, such as feature interdependencies.

### *Generative models*

Generative modeling relies on the use of statistics and probability to extract patterns from data and learn the underlying distribution. In the following, three types of generative integrative AD models are reviewed: event-based models, Bayesian network learning, and autoencoders.

*Event-based models.* Event-based models estimate the most probable sequence of events based on the assessment of a probability density function for a particular event order. Fonteijn et al. (2012), Chen et al. (2016), and Oxtoby et al. (2018), used this method to learn the sequence of AD events based on imaging and non-imaging measurements from a clinical study. The authors first fitted simple mixture models (e.g., gaussian mixture models) to individual biomarkers in order to calculate the likelihood of the normality or abnormality status per biomarker. Given these likelihoods, by multiplication of the probabilities, the likelihoods for each possible order of events was calculated. The order with the highest probability was then selected using a greedy Markov Chain Monte Carlo algorithm to describe the temporal correlation of the biomarker trajectories over the course of AD progression.

The models developed by Fonteijn et al. (2012) and Chen et al. (2016) simplified the sequence of biomarker abnormalities over the course of the disease progression by relying on the assumption that all subjects follow a single event sequence. However, AD is highly heterogeneous and includes distinct subgroups (Ferreira et al., 2018). To account for this, Young et al. (2015) established their event-based models with two extensions: a Mallows model and a Dirichlet process mixture of generalized Mallows models. The first extension allows subjects to deviate from the main event sequence, and the latter clusters subjects according to different event sequences.

In principle, the event sequence proposed in the hypothetical model is similar to that observed using traditional and event-based models. Changes in CSF measures are the earliest events, followed by regional brain atrophies and finally succeeded by diminished cognitive scores. However, the event sequence in the hypothetical and traditional models is constructed based on predefined clinical assessments and often imprecise or subjective cut-offs. By contrast, in generative models, the sequence of events, as well as the clustering of biomarkers into normal and abnormal classes, is directly extracted from the data (e.g., the onset of a new symptom, like memory performance decline). Thus, event-based models explain the changes without a priori

biases. Moreover, generative models are able to characterize uncertainty in the event ordering arising from heterogeneity in the population and thus, can address individual deviations from the generic model.

*Bayesian network learning.* Extensive research efforts have been made to uncover the relationships between individual biomarkers and AD. Yet the number of studies that investigated the interplay between multiple biomarkers themselves is comparably limited. Khanna et al. (2018) and Ding et al. (2018) built Bayesian network models covering different biological scales and time points to uncover the interplay amongst sets of biomarkers. Ding et al. (2018) considered the ApoE allele, PET and MRI imaging data, scores from psychological and functional tests, and the medical history of patients with respect to neurological diseases. Using a variety of feature selection metrics, they determined the most relevant features with respect to the clinical dementia rating and modeled these heterogeneous measurements using a Bayesian network to determine their probabilistic interdependencies. However, these models only capture conditional probabilities between predictor variables and clinical outcomes. They are unable to provide a causal mechanistic understanding of an observed phenomenon. Such hypothesized pathophysiological mechanisms are important for making reliable predictions and having confidence in the practical application of data-driven models. To this end, Khanna et al. (2018) employed a combination of data-driven probabilistic and knowledge-driven mechanistic approaches. They modeled clinical variables, genetic variants, pathways, and neuro-imaging readouts using Bayesian network learning to estimate dependencies between disease relevant features. Together with a cause-and-effect knowledge model derived from scientific literature, they partially reconstructed biological mechanisms that could play a role in the conversion of normal/MCI into AD pathology.

*Autoencoders.* The last type of generative model discussed in this review is autoencoders. In essence, an autoencoder is a neural network that aims to encode the input data into a lower dimensional representation and from that decode it again, reconstructing the original input. It has successfully been applied for different tasks on AD cohorts (Basu et al., 2019; Martinez-Murcia et al., 2019). The two main applications of this approach in the field consist of classifying patients based on AD diagnosis (Basu et al., 2019) and clustering of patient trajectories into subgroups (De Jong et al., 2019). These strategies are especially interesting for patient classification and stratification tasks in datasets where information is sparse. However, another novel and promising task for autoencoders is the generation of synthetic data from real patient level data (Gootjes-Dreesbach et al., 2019). This, in turn, could be used to circumvent legal and ethical constraints that restrict data sharing.

### *Discriminative models*

Discriminative models are a class of models generally used for classification. Discriminative models that rely on labeled data are called supervised models, while unsupervised models use unlabeled data.

*Supervised discriminative models.* Diverse supervised discriminative methods such as support vector machines (SVM; Magnin et al., 2010), and multiple-kernel SVM (MKL; Hinrichs et al., 2010; Zhang et al., 2011) have been used to classify AD patients, MCI subjects, and controls. However, studies that used multiple-kernel SVM reported superior classification performance, because the use of multiple kernels facilitates the integration of multimodal biomarker data (Zhang et al., 2011). Additionally, MKL are well-suited for dealing with very high dimensional data (Young et al., 2013). MKL also enable individual weighting of biomarker modalities. This offers more flexibility for kernel combination and thus, a better integration of the data. For example Hinrichs et al. (2010), applied MKL in combination with MRI and PET imaging to differentiate between AD subjects and controls. Their method showed high classification performance, achieving 92.4% accuracy. Similarly, Zhang et al. (2018) combined MRI, PET, and CSF biomarkers to discriminate between healthy controls and AD/MCI. After integrating all biomarker data using a MKL, they deployed a linear SVM for the actual classification task, which resulted in 93.2% accuracy for classifying AD and healthy controls and 76.4% for discriminating between MCI and healthy controls. Both studies applied a similar method for classification, yet the latter one achieved a slightly higher accuracy. Comparing the approaches applied in Zhang et al. (2018) and Hinrichs et al. (2010) it becomes clear that the major reason for the difference in performance is the feature selection process. Depending on the available sample size, other methods might prove more promising (Liu et al., 2012). Moreover, Zhang et al. (2018) benefits from employing three biomarker modalities, namely, CSF measurements and two imaging modalities, compared to Hinrichs et al. (2010) who only use the two imaging modalities.

While the above kernel-based pattern recognition approaches yield categorical class decisions, Young et al. (2013) used gaussian process classification, which is a probabilistic classification algorithm. This study integrated imaging, CSF, neuropsychological, and genetic biomarkers to classify MCI subjects who remained stable and MCI patients who converted to AD within 3 years. In contrast to MKL, the probabilistic classification afforded by the gaussian process approach provides the opportunity to position the subjects according to disease stage, to stratify patients, and to model the sequence order of biomarker abnormality.

Another type of discriminative model is disease risk models. This type of supervised model can be used to predict the time to AD diagnosis for normal/MCI patients. Multiple approaches have been used to develop risk models for AD (Da et al., 2013; Li et al., 2013). Li et al. (2013) used a combination of cox regression analyses and time-dependent ROC approaches to evaluate prognostic utility and performance stability of candidate biomarkers. The authors deduced that both baseline volumetric MRI and cognitive measures can predict progression from MCI to AD. However, in participants' follow-up visits, only cognitive measurements remained predictive. Da et al. (2013) employed

the cox proportional hazards models to compare the magnitudes of the relative association between predictors (patterns of brain atrophy, cognitive assessments, genetics, and CSF biomarkers) and time to conversion from MCI to AD. They concluded that brain atrophy and cognitive assessments in combination offer the highest predictive power of conversion from MCI to AD.

Although the results in both studies were similar, the time-dependent ROC curve used by Li et al. (2013) enabled them to predict disease risk as a function of time. Thus, this method provides clear benefit for a progressive disease such as AD, in which both the disease status and biomarker measurements change over time (Kamarudin et al., 2017).

The data labeling which enables supervised discriminative models to determine decision boundaries for distinguishing classes of interest can also introduce errors. Inaccurate labels will negatively affect the performance of the classifier. Such mislabeling is not uncommon in AD, due to the absence of a clear diagnostic biomarker (Fischer et al., 2017). Instead, diagnosis is currently made based on symptoms (Schott and Petersen, 2015) Furthermore, integrative data analysis is further complicated by the fact that the diagnostic criteria for MCI have changed over the years, and MCI is not consistently defined across clinical studies. While one study relies on assessing only a single cognitive domain for MCI diagnosis, such as speech or memory, others base their diagnoses on performance on cognitive tests for multiple domains. Apart from that, there are multiple pathologies for MCI; AD is just one of them. Thus, unified clear disease definitions are crucial, since the MCI classification accuracy can influence outcomes of research and clinical practice (Jak et al., 2010).

*Unsupervised Discriminative Models.* Unsupervised discriminative models use a variety of clustering techniques on unlabeled data, avoiding the challenges of data label accuracy. These techniques use properties of each data point to iteratively form groups, called clusters. This ultimately leads to a discrimination of the data into several clusters of highly similar data points. Given the observed biological heterogeneity among normal control subjects, Nettiksimmons et al. (2014) hypothesized that different subgroups may also be found among the MCI subjects. Using agglomerative hierarchical clustering, they sorted subjects based on MRI volumes, CSF measurements, and cognitive tests. Next, the resulting clusters were explored with regard to longitudinal atrophy, conversion time, and cognitive trajectories. Four clusters with unique biomarker patterns resulted: (i) a cluster biologically similar to normal controls. MCI patients from that cluster rarely converted to AD, (ii) one cluster with early AD pathology characteristics, (iii) another cluster of subjects with hardly any tau abnormality, but a high proportion of AD converters, and (iv) and finally one cluster with pre-AD symptoms wherein almost all subjects converted to AD. Based on these findings, they hypothesized that clusters ii and iv reflected the amyloid cascade pattern (Ricciarelli and Fedele, 2017) since both clusters presented lower CSF Aβ levels and elevated tau proteins. However, the tau level in cluster iv was higher, and more severe atrophy as well as cognitive impairment were detected. The authors concluded that more tau accumulation may lead to more cognitive decline. One of the intrinsic limitations of their clustering approach is that the number of clusters must be predefined. The maximum gap statistic is one approach to determine this number (Tibshirani et al., 2001). However, specifying the number of clusters beforehand will always bias the clustering to some extent, and choosing a reasonable number is no trivial task given the broad variety of subtypes found among AD subjects.

Toschi et al. (2019) used Density-Based Spatial Clustering of Applications with Noise (DBSCAN; Thanh et al., 2013), which does not require pre-specifying the number of clusters. They integrated five validated CSF biomarkers in order to cluster a cohort where symptomatic patients presented diagnoses ranging from self-perceived cognitive decline (Zhang et al., 2011) to MCI to AD. In contrast to the previous study, Toschi et al. (2019) adjusted all biomarker values for age, sex and their interactions to exclude them as confounders (Pourhoseingholi et al., 2012). Moreover, Toschi et al. (2019) used t-Distributed Stochastic Neighbor Embedding (t-SNE) to reduce the dimensionality of biomarkers space, since defining the distance between the data points in a high dimensional space of biomarkers is notoriously difficult (Domingos, 2012). Finally, they applied DBSCAN on this lower dimensional representation. DBSCAN defines a high data density region based on two parameters: (i) the radius of the neighborhood, and (ii) the minimum number of points within the radius. These values are determined by a nearest neighbor method, in which the distance of each point to their nearest n points is calculated. Afterwards, results are sorted, plotted and the value with most pronounced change is selected as the optimal value. Using DBSCAN, Toschi et al. (2019) characterized five biological clusters which were not significantly bound to the original distinct clinically phenotyped diagnostic groups. They explained that the clusters included all phenotypic groups and were not homogeneous enough to be considered as a specific AD pathophysiology. Moreover, contrary to general belief that $A\beta_{1-42}$ is linearly associated with the progression of AD and cognitive decline (Sperling et al., 2011a; Samtani et al., 2013), their findings suggest that $A\beta_{1-42}$ is less likely to contribute to phenotypic discrimination.

The dimensionality reduction technique, t-SNE, used by Toschi et al. (2019) enabled them to better separate the data and hence, to enhance cluster identification, in comparison to directly running a clustering algorithm on a high dimensional data as Nettiksimmons et al. (2014). However, their main limitation is that clustering results are highly sensitive to two parameters necessary for DBSCAN. Moreover, they did not include other biomarkers, such as imaging and genetics biomarkers, which could enhance their clustering, as previously reported by Young et al. (2013, 2018).

Unsupervised clustering algorithms are ideal for identifying subgroups and non-linear associations between individuals based on a multidimensional profile, regardless of the individual labels, in contrast to supervised algorithms. This allows the grouping of individuals based on shared pathophysiological drivers and triggers and, possibly, similar longitudinal disease trajectories. This is an advantage in the AD field due to the prevalence of unreliable labels stemming from misdiagnosis

and to the biological heterogeneity of AD subjects. On the other hand, most unsupervised clustering algorithms perform better with a larger sample size than is often obtainable in AD studies (Oxtoby and Alexander, 2017). Therefore, the smaller size inherent to AD cohorts may lead to clustering instability.

To this point, we have reviewed a broad variety of data-driven integrative AD models and elaborated on their associated limitations and challenges. In the following, we enumerate more general challenges researchers encounter in the course of data-driven integrative AD modeling and suggest how these could be addressed.

## Challenges of Data-Driven Modeling

Although there exists a wide range of data-driven integrative modeling approaches, not all of them are well-suited for every analytic task and each has its own strengths and weaknesses. Still, there are some challenges which affect all data-driven approaches to some degree: data collection, reproducibility of findings, and interpretability of models and results.

### Data Collection

Collecting patient level data, the basis for all data-driven modeling, is a time-consuming and costly process. Additionally, it is a source of major challenges and limitations of these models. In particular, data "censoring" and "missingness," can impede modeling, bias models, or even make certain modeling techniques unfeasible.

Data censoring describes the condition in which a particular event (here AD diagnosis) is not observed for certain study participants during the study runtime. This censoring can occur in two ways: if AD diagnosis occurred before the start of the study; or if the patient drops out of the study, or the study ends without occurrence of the AD diagnosis event. A significant number of patients enrolled in clinical studies have already received a diagnosis before the beginning of the study, indicating that they are in a progressed stage of the disease (Ellis et al., 2009). It is therefore not possible to obtain indications of early disease onset in such patients. The second form of censoring arises from two sources. First, all observational cohort studies experience participant dropout for a variety of reasons, including the participation burden on caregivers or medical problems (Coley et al., 2008). Second, subjects that remain healthy throughout study runtime could still develop the disease after the study ended, meaning they were in a prodromal disease stage. It is thus impossible to know if or when the patient would eventually receive an AD diagnosis. This form of censoring is common in longitudinal AD studies, because AD is a slow-progressing disease, while the studies are typically quite short (Lawrence et al., 2017), due to limited funding (Prabhakaran and Bakshi, 2018).

Disease onset is a critical point for clinical intervention (Sperling et al., 2011b), so it is subject to extensive research efforts. It is here, however, where data censoring impedes data analysis the most. Data censoring can result in over- or under-sampling of early and advanced disease stages. This, in turn, leads to models biased toward specific disease stages (Ning et al., 2010). Various methods, such as complete data analysis (Xiang et al., 2013), imputation (Fisher et al., 2019), or analysis based on dichotomized data (Donohue et al., 2011), have been established to address censored data. Yet all of these methods may introduce error and impose complexities and biases on other integrative modeling steps, such as model interpretation, and thus need to be used with care (Prinja et al., 2010).

The complete absence of a value for variables in the observation of interest likewise poses a significant challenge to data-driven modeling. This missing data in AD cohort studies occurs for several reasons, including unwillingness of patients to undergo invasive tests like lumbar punctures, and the high cost of measuring a particular variable, such as imaging scans (Engelborghs et al., 2017). The implications of such a scenario include a loss of statistical power of the study and may bias the conclusions that can be drawn (Hughes et al., 2019). Over the past decades, novel statistical methods (Molenberghs et al., 2014) and software (Quartagno and Carpenter, 2016; Moreno-Betancur et al., 2017) have been developed for analyzing data with missing values. However, analysis restricted to individuals with complete data is generally preferred, if feasible.

Despite the challenges in collecting complete and uncensored data, the value of data in strengthening disease understanding is clear. Several large-scale AD patient datasets have been collected for use in a variety of studies (Lawrence et al., 2017) including, for example, Alzheimer's Disease Neuroimaging Initiative (ADNI; Mueller et al., 2005), Australian Imaging Biomarkers and Lifestyle Study of Aging (AIBL; Ellis et al., 2009), the Dominantly Inherited Alzheimer Network (DIAN; Moulder et al., 2013), and European Prevention of Alzheimer's Dementia (EPAD; Vermunt et al., 2018). However, these classical observational studies are subject to bias, resulting from the inclusion and exclusion criteria used to select participants (Miksad and Abernethy, 2018).

The use of electronic medical records (EMRs) has been proposed as a potential solution to reduce the bias of classical clinical trials. They provide an alternative view on patient measurements (Fröhlich et al., 2018), so, a collection of EMRs can provide a more representative view on patient measurements. However, EMRs are largely phenotypic: molecular phenomena such as genomic variants are not reflected in the data. Moreover, extracting information from EMRs requires natural language preprocessing, which itself currently remains a difficult and error-prone process.

### Reproducibility

The ability to reproduce the findings of a study using different subjects is an important part of scientific research. This is particularly the case in integrative AD modeling, since the tendency of AD datasets is not to fully reflect the diversity of AD patients. Inclusion-exclusion criteria in clinical studies can lead to significant under-representation of some populations. For example, the landscape of data-driven AD models is currently dominated by only a few cohorts which are made up largely of White Caucasians, and, to a lesser extent, are constrained by geographic location (Lawrence et al., 2017). Since most observational cohorts are not representative of the general AD population (Ferreira et al., 2017), it is important to validate the resulting models with an independent cohort study. While this

external validation is a necessary step to corroborate findings, it is complicated by data interoperability and sample size.

### *Interoperability*

The ability to map the data coming from one study to data from another study is known as data interoperability[5]. Each of the major AD clinical studies was established with a specific sample and feature characterization. Since they might not be directly interoperable, extensive curation is needed before the external validation of a model can be carried out. Otherwise, the training cohort and the validation cohort would be based on different populations, and would contain different measurements. Thus, before validation, researchers must map and assess the "comparability" of both features and subjects.

Feature mapping requires specifying relationships between data elements from different data models and standardizing the terms used to represent the features in the two datasets. This is due to the fact that controlled vocabularies are not used to annotate the datasets. Thus, even if the same biomarker has been collected in two studies, it is usually referred to by different terms, impeding a direct comparison of the datasets. For example, the hippocampus is one of the earliest sites of AD pathology, and hippocampal volume is measured in ADNI and EPAD. However, ADNI identifies this biomarker as "Hippocampus," while EPAD refers to it as "lhvr" (right hemisphere) and "lhvl" (left hemisphere).

Moreover, the subject populations in each study must be comparable. For instance, if the biological sex distributions in two AD studies differ significantly, then the cognitive impairment scores of the cohorts cannot be directly compared, because female AD patients have been shown to have greater cognitive impairment than men in comparable stages of the disease (Laws et al., 2016).

There are several strategies to overcome the lack of interoperability between datasets at both feature and subject level. At the feature level, interoperability can be attained by annotating datasets according to a standard controlled vocabulary. Several such vocabularies (e.g., NIFT Iyappan et al., 2017 and PTS Iyappan et al., 2016) have been established, but significant improvements in interoperability will only come with widespread adoption (Neu et al., 2012). The most prominent example might be the AD specific standard developed by the Clinical Data Interchange Standards Consortium (CDISC; Neville et al., 2017). At the subject level, mapping between training and validation cohorts can be accomplished by identifying, in the validation cohort, a subset of subjects that is statistically comparable to the training cohort. Finally, in order to assess the comparability of subjects from different studies, techniques such as statistical matching can be used (Austin, 2011).

### *Sample size*

The relatively small sample sizes of AD clinical studies also contributes to the challenge of reproducibility in AD integrative modeling. Many AD studies contain fewer than a thousand patients, and the longitudinal follow-up is limited. In addition, typically not all of the subjects were screened for the complete biomarker set, leading to sparse subsets of patients for whom the study contains complete data. As a result, models generated from these studies have a high margin of error and low statistical power, meaning they struggle to detect small effects.

The integration of different datasets into a larger dataset can overcome some of the challenges related to small sample sizes (Gomez-Cabrero et al., 2014). Integrated datasets provide more comprehensive data, and the resulting models have greater statistical power. However, current approaches for data integration were developed for the analysis of single-data-type datasets, and only subsequently adapted to handle datasets with multiple data types. For this reason, data integration methodologies can be ill-suited to manage the computational challenges arising from the variety of different data sizes, formats, and dimensionalities present in AD datasets, as well as their noisiness, complexity, and the level of agreement between datasets (Gomez-Cabrero et al., 2014; Gligorijević et al., 2015). Furthermore, even data acquired by analogous technologies are not necessarily integrable. For example, neuroimaging data acquired from similar scanners and similar modalities may still be stored in different formats and have different metadata content (Goble and Stevens, 2008).

Several strategies could be applied to address the interoperability challenges arising from data integration. The first strategy is to normalize and standardize data across all platforms (O'Bryant et al., 2015). However, scientific independency and freedom for innovation, as well as uniqueness of databases, must be respected. The second strategy is to collect a standardized set of biomarkers across different studies. Finally, the ideal solution would be performing a systematic longitudinal clinical and -omics follow-up of each individual in a large and rigorously characterized cohort since this would provide a statistically sufficient number of measurements in the context of subjects and variables. The Deep and Frequent Phenotyping study from Lawson et al. (2017) showed that such a cohort, in theory, is feasible. Yet, including a sufficient number of participants in such an ambitious study is costly.

## Interpretability

In order for an AD model to have clinical impact, its findings must be interpretable. There are several barriers to AD model interpretability. Machine learning models often act as "black boxes"; it may be impossible to uncover the reasons for the predictions made by the model (Rudin, 2019). Indeed, as the number of features and the complexity of the computational processes used in models increases, this interpretability problem will worsen. Moreover, data-driven models are not causal and typically capture non-linear correlations between predictor and explanatory variables. While prior understanding of cause–effect relationships and detailed mechanisms might prove helpful to well-performing models, it is not necessarily required. Lack of mechanistic explanations for model prediction complicates the interpretation of data-driven findings and reduces acceptance by physicians (Fröhlich et al., 2018). Thus, the translation of data-driven models into a biomedical knowledge context is a major challenge in integrative AD modeling.

[5]https://library.ahima.org/doc?oid=65895#.Xdl-iZPYrOQ

Combining available mechanistic knowledge with machine learning-based sub-models, so-called hybrid modeling could bridge the gap between experimental biological and computational research by improving interpretability (Fröhlich et al., 2018). For example, Bayesian networks which built on causal knowledge graphs constitute such a hybrid model (Arora et al., 2019). They shed light on interdependencies across features, which can be on different scales (e.g., clinical, genetic, and molecular), and allow for predicting the outcome of purely hypothetical clinical interventions. Similarly, other recent deep learning methodologies use knowledge-derived biological networks to define the layers of neural networks in order to improve interpretability (Fortelny and Bock, 2019).

## CONCLUSION

In the era of extensive biomarker profiling, big data, and artificial intelligence, integrative AD modeling comes with high promises. By integrating multi-scale, multimodal, and longitudinal patient data, such modeling approaches aim to provide a holistic picture of disease pathophysiology and progression. However, as we have discussed in this review, while integrative models have generated significant insights, and thus proved to be valuable in research, existing models do not yet fully describe critical aspects of AD.

The construction of hypothetical models simultaneously benefits and suffers from the vast amount of published knowledge. Prioritization of articles and computational text mining of literature corpora are reasonable approaches to identify a greater quantity of relevant knowledge while designing hypothetical models. In the field of data-driven integrative AD modeling, we highlighted several major ongoing challenges throughout the whole modeling process of data collection, integration of disparate data sources, data analysis, and model interpretation. Data missingness and data censoring are major bottlenecks in data collection as well as analysis and interpretation. Heterogeneity and complexity in biological data are major impediments to data integration, which is central to data-driven integrative modeling and validation. Data mapping, imprecise diagnostic stages, and biased data are barriers that hamper data analysis and interpretation. Furthermore, there is an insufficient number of subjects in studies, which restricts the statistical power of data-driven integrative AD models. Because of these challenges, to the best of our knowledge, at this point in time, there are no integrative AD models which have been used in clinical practice.

While in theory, certain existing integrative models are capable of predicting AD diagnosis and progression, they are not used in clinical practice. We see a number of steps that could bring us closer to the goal of precision medicine and that could enable patient diagnosis through integrative disease models in a clinical context. First, we, the AD research community, need to establish valid, informative biomarkers and clear criteria for AD diagnosis. This would result in reliable predictors that could be included in modeling approaches, as well as fewer diagnostic errors, which in turn reduce the effect of mislabeled data. Second, a global data schema that could support the normalization and standardization of data across measurements would ultimately facilitate improved data integration. If future cohort studies would adhere to such a schema, data integration would be straightforward and the cumulative time saved for researchers working with it would be enormous. Finally, innovative modeling approaches, such as causal inference techniques and hybrid modeling, which go beyond current state-of-the-art data-driven models by linking prior knowledge with data-driven models, need to be developed and made more robust. Overall, novel computational modeling approaches that surmount the current integrative AD modeling challenges may hold the potential to play an increasing role in the planning of medical interventions and practice.

## AUTHOR CONTRIBUTIONS

SG drafted the manuscript. CR, CB, DD-F contributed to the final version of the manuscript. CH and MH-A reviewed the final version of the manuscript.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the B-IT foundation.

## REFERENCES

Aisen, P. S., Cummings, J., Jack, C. R. Jr., Morris, J. C., Sperling, R., Frölich, L., et al. (2017). On the path to 2025: understanding Alzheimer's disease continuum. *Alzheimers Res. Ther.* 9:60. doi: 10.1186/s13195-017-0283-5

Anne, M., and Fagan. (2014). CSF biomarkers of Alzheimer's disease: impact on disease concept, diagnosis, and clinical trial design. *Adv. Geriatr.* 2014:302712. doi: 10.1155/2014/302712

Arora, P., Boyne, D., Slater, J. J., Gupta, A., Brenner, D. R., and Druzdzel, M. J. (2019). Bayesian networks for risk prediction using real-world data: a tool for precision medicine. *Val. Health* 22, 439–445. doi: 10.1016/j.jval.2019.01.006

Atkins, S., Clear, J., and Ostler, N. (1992). Corpus design criteria. *Lit. Ling. Comput.* 7, 1–16. doi: 10.1093/llc/7.1.1

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.* 46, 399–424. doi: 10.1080/00273171.2011.568786

Bartlett, J. W., Frost, C., Mattsson, N., Skillbäck, T., Blennow, K., Zetterberg, H., et al. (2012). Determining cut-points for Alzheimer's disease biomarkers: statistical issues, methods and challenges. *Biomark. Med.* 6, 391–400. doi: 10.2217/bmm.12.49

Basu, S., Wagstyl, K., Zandifar, A., Collins, L., Romero, A., and Precup, D. (2019). "Early prediction of alzheimer's disease progression using variational autoencoders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, eds D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan (Cham: Springer), 205–213. doi: 10.1007/978-3-030-32251-9_23

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., et al. (2012). Clinical and biomarker changes in dominantly inherited

Alzheimer's disease. *N. Engl. J. Med.* 367, 795–804. doi: 10.1056/NEJMoa1202753

Blennow, K., and Zetterberg, H. (2018). Biomarkers for Alzheimer's disease: current status and prospects for the future. *J. Intern. Med.* 284, 643–663. doi: 10.1111/joim.12816

Boutron, I., and Ravaud, P. (2018). Misrepresentation and distortion of research in biomedical literature. *Proc. Natl. Acad. Sci. U.S.A.* 115, 2613–2619. doi: 10.1073/pnas.1710755115

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Front. Neurosci.* 11:543. doi: 10.3389/fnins.2017.00543

Caroli, A., and Frisoni, G. B. (2010). The dynamics of Alzheimer's disease biomarkers in the Alzheimer's disease neuroimaging initiative cohort. *Neurobiol. Aging* 31, 1263–1274. doi: 10.1016/j.neurobiolaging.2010.04.024

Chen, G., Shu, H., Chen, G., Ward, B. D., Antuono, P. G., Zhang, Z., et al. (2016). Staging Alzheimer's disease risk by sequencing brain function and structure, cerebrospinal fluid, and cognition biomarkers. *J. Alzheimers Dis.* 54, 983–993. doi: 10.3233/JAD-160537

Chuang, Y. F., An, Y., Bilgel, M., Wong, D. F., Troncoso, J. C., O'Brien, R. J., et al. (2016). Midlife adiposity predicts earlier onset of Alzheimer's dementia, neuropathology and presymptomatic cerebral amyloid accumulation. *Mol. Psychiatry* 21, 910–915. doi: 10.1038/mp.2015.129

Coley, N., Gardette, V., Toulza, O., Gillette-Guyonnet, S., Cantet, C., Nourhashemi, F., et al. (2008). Predictive factors of attrition in a cohort of Alzheimer disease patients. *Neuroepidemiology* 31, 69–79. doi: 10.1159/000144087

Da, X., Toledo, J. B., Zee, J., Wolk, D. A., Xie, S. X., and Ou, Y. (2013). Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. *Neuroimage Clin.* 4, 164–173. doi: 10.1016/j.nicl.2013.11.010

De Jong, J., Emon, M. A., Wu, P., Karki, R., Sood, M., Godard, P., et al. (2019). Deep learning for clustering of multivariate clinical patient trajectories with missing values. *GigaScience* 8:giz134. doi: 10.1093/gigascience/giz134

Ding, X., Bucholc, M., Wang, H., Glass, D. H., Wang, H., Clarke, D. H., et al. (2018). A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. *Sci. Rep.* 8:9774. doi: 10.1038/s41598-018-27997-8

Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM* 55, 78–87. doi: 10.1145/2347736.2347755

Dong, R., Wang, H., Ye, J., Wang, M., and Bi, Y. (2019). Publication trends for Alzheimer's disease worldwide and in China: a 30-year bibliometric analysis. *Front. Hum. Neurosci.* 13:259. doi: 10.3389/fnhum.2019.00259

Donohue, M. C., Gamst, A. C., Thomas, R. G., Xu, R., Beckett, L., Petersen, R. C., et al. (2011). The relative efficiency of time-to-threshold and rate of change in longitudinal data. *Contemp. Clin. Trials* 32, 685–693. doi: 10.1016/j.cct.2011.04.007

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., et al. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int. Psychogeriatr.* 21, 672–687. doi: 10.1017/S1041610209009405

Engelborghs, S., Niemantsverdriet, E., Struyfs, H., Blennow, K., Brouns, R., Comabella, M., et al. (2017). Consensus guidelines for lumbar puncture in patients with neurological diseases. *Alzheimers Dement.* 8:111–126. doi: 10.1016/j.dadm.2017.04.007

Ferreira, D., Hansson, O., Barroso, J., Molina, Y., Machado, A., Hernández-Cabrera, J. A., et al. (2017). The interactive effect of demographic and clinical factors on hippocampal volume: a multicohort study on 1958 cognitively normal individuals. *Hippocampus* 27, 653–667. doi: 10.1002/hipo.22721

Ferreira, D., Wahlund, L., and Westman, E. (2018). The heterogeneity within Alzheimer's disease. *Aging* 10, 3058–3060. doi: 10.18632/aging.101638

Fischer, C. E., Qian, W., Schweizer, T. A., Ismail, Z., Smith, E. E., Millikin, C. P., et al. (2017). Determining the impact of psychosis on rates of false-positive and false-negative diagnosis in Alzheimer's disease. *Alzheimers Dement.* 3, 385–392. doi: 10.1016/j.trci.2017.06.001

Fisher, C. K., Smith, A. M., and Walsh, J. R. (2019). Machine learning for comprehensive forecasting of Alzheimer's disease progression. *Sci. Rep.* 9:13622. doi: 10.1038/s41598-019-49656-2

Fonteijn, H. M., Modat, M., Clarkson, M. J., Barnes, J., Lehmann, M., Hobbs, N. Z., et al. (2012). An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *Neuroimage* 60, 1880–1889. doi: 10.1016/j.neuroimage.2012.01.062

Fortelny, N., and Bock, C. (2019). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. doi: 10.1101/794503

Frisoni, G. B., Fox, N. C., Jack, C. R. Jr, Scheltens, P., and Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* 6, 67–77. doi: 10.1038/nrneurol.2009.215

Fröhlich, H., Balling, R., Beerenwinkel, N., Kohlbacher, O., Kumar, S., Lengauer, T., et al. (2018). From hype to reality: data science enabling personalized medicine. *BMC Med.* 16:150. doi: 10.1186/s12916-018-1122-7

Gamberger, D., Lavrač N., Srivatsa, S., Tanzi, R. E., and Doraiswamy, P. M. (2017). Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. *Sci. Rep.* 7:6763. doi: 10.1038/s41598-017-06624-y

Gladun, V. P. (1997). Hypothetical modeling: methodology and application. *Cybern. Syst. Anal.* 33, 7–15. doi: 10.1007/BF02665935

Gligorijević V., and PrŽulj, N. (2015). Methods for biological data integration: perspectives and challenges. *J. R. Soc. Interface* 12:20150571. doi: 10.1098/rsif.2015.0571

Goble, C., and Stevens, R. (2008). State of the nation in data integration for bioinformatics. *J. Biomed. Inform.* 41, 687–693. doi: 10.1016/j.jbi.2008.01.008

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., et al. (2014). Data integration in the era of omics: current and future challenges. *BMC Syst. Biol.* 8(Suppl. 2):I1. doi: 10.1186/1752-0509-8-S2-I1

Gootjes-Dreesbach, L., Sood, M., Sahay, A., Hofmann-Apitius, M., and Fröhlich, H. (2019). Variational Autoencoder Modular Bayesian Networks (VAMBN) for simulation of heterogeneous clinical study data. *bioRxiv* 760744. doi: 10.1101/760744

Gyori, B. M., Bachman, J. A., Subramanian, K., Muhlich, J. L., Galescu, L., and Sorger, P. K. (2017). From word models to executable models of signaling networks using automated assembly. *Mol. Syst. Biol.* 13:954. doi: 10.15252/msb.20177651

Hampel, H., O'Bryant, S. E., Durrleman, S., Younesi, E., Rojkova, K., Escott-Price, V., et al. (2017). A precision medicine initiative for Alzheimer's disease: the road ahead to biomarker-guided integrative disease modeling. *Climacteric* 20, 107–118. doi: 10.1080/13697137.2017.1287866

Hinrichs, C., Singh, V., Xu, G., and Johnson, S. (2010). MKL for robust multi-modality AD classification. *Med. Image Comput. Comput. Assist. Interv.* 12(Pt 2), 786–794. doi: 10.1007/978-3-642-04271-3_95

Hinrichs, C., Singh, V., Xu, G., and Johnson, S. C. (2011). Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *Neuroimage* 55, 574–589. doi: 10.1016/j.neuroimage.2010.10.081

Hughes, R. A., Heron, J., Sterne, J. A. C., and Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *Int. J. Epidemiol.* 48, 1294–1304. doi: 10.1093/ije/dyz032

Humayun, F., Domingo-Fernandez, D., George, A. A. P., Hopp, M. T., Syllwasschy, B. F., Detzel, M., et al. (2019). A computational approach for mapping heme biology in the context of hemolytic disorders. *bioRxiv* 804906. doi: 10.1101/804906

Iyappan, A., Gündel, M., Shahid, M., Wang, J., Li, H., Mevissen, H. T., et al. (2016). Towards a pathway inventory of the human brain for modeling disease mechanisms underlying neurodegeneration. *J. Alzheimers Dis.* 52, 1343–1360. doi: 10.3233/JAD-151178

Iyappan, A., Younesi, E., Redolfi, A., Vrooman, H., Khanna, S., Frisoni, G. B., et al. (2017). Neuroimaging feature terminology: a controlled terminology for the annotation of brain imaging features. *J. Alzheimers Dis.* 59, 1153–1169. doi: 10.3233/JAD-161148

Jack, C. R. Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., et al. (2016a). A/T/N: an unbiased descriptive

classification scheme for Alzheimer disease biomarkers. *Neurology* 2, 539–547. doi: 10.1212/WNL.0000000000002923

Jack, C. R. Jr., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., et al. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* 12, 207–216. doi: 10.1016/S1474-4422(12)70291-0

Jack, C. R. Jr., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., et al. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* 9, 119–128. doi: 10.1016/S1474-4422(09)70299-6

Jack, C. R. Jr., Vemuri, P., Wiste, H. J., Weigand, S. D., Aisen, P. S., Trojanowski, J. Q., et al. (2011). Evidence for ordering of Alzheimer disease biomarkers. *Arch. Neurol.* 68, 1526–1535. doi: 10.1001/archneurol.2011.183

Jack, C. R. Jr., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Lowe, V., et al. (2012). Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. *Arch. Neurol.* 69, 856–867. doi: 10.1001/archneurol.2011.3405

Jack, C. R. Jr., Wiste, H. J., Weigand, S. D., Therneau, T. M., Lowe, V. J., and Knopman, D. S. (2016b). Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement.* 13, 205–216. doi: 10.1016/j.jalz.2016.08.005

Jak, A. J., Bondi, M. W., Delano-Wood, L., Wierenga, C., Corey-Bloom, J., Salmon, D. P., et al. (2010). Quantification of five neuropsychological approaches to defining mild cognitive impairment. *Am. J. Geriatr. Psychiatry* 17, 368–375. doi: 10.1097/JGP.0b013e31819431d5

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51, 404–413. doi: 10.1038/s41588-018-0311-9

Kamarudin, A. N., Cox, T., and Kolamunnage-Dona, R. (2017). Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med. Res. Methodol.* 17:53. doi: 10.1186/s12874-017-0332-6

Khanna, S., Domingo-Fernández, D., Iyappan, A., Emon, M. A., Hofmann-Apitius, M., and Fröhlich, H. (2018). Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's disease and reconstruction of relevant biological mechanisms. *Sci. Rep.* 8:11173. doi: 10.1038/s41598-018-29433-3

Klunk, W., Cohen, A., Bi, W., Weissfeld L,Aizenstein, H., McDade, E., et al. (2012). Why we need two cutoffs for amyloid imaging: early versus Alzheimer's-like amyloid-positivity. *Alzheimers Dement.* 8, P453–P454. doi: 10.1016/j.jalz.2012.05.1208

Kudelic R., Konecki, M., and Malekovic, M. (2011). "Mind map generator software model with text mining algorithm," in *Proceedings of the ITI 2011, 33rd International Conference on Information Technology Interfaces* (Cavtat), 487–494.

Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., et al. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* 51, 414–430. doi: 10.1038/s41588-019-0358-2

Lamurias, A., and Couto, F. M. (2019). "Text mining for bioinformatics using biomedical literature," in *Encyclopedia of Bioinformatics and Computational Biology*, eds S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach (Oxford: Elsevier), 602–611. doi: 10.1016/B978-0-12-809633-8.20409-3

Lawrence, E., Vegvari, C., Ower, A., Hadjichrysanthou, C., De Wolf, F., and Anderson, R. M. (2017). A systematic review of longitudinal studies which measure alzheimer's disease biomarkers. *J. Alzheimers Dis.* 59, 1359–1379. doi: 10.3233/JAD-170261

Laws, K. R., Irvine, K., and Gale, T. M. (2016). Sex differences in cognitive impairment in Alzheimer's disease. *World J. Psychiatry* 6, 54–65. doi: 10.5498/wjp.v6.i1.54

Lawson, J., Murray, M., Zamboni, G., Koychev, I. G., Ritchie, C. W., Ridha, B. H., et al. (2017). Deep and frequent phenotyping: a feasibility study for experimental medicine in dementia. *J Alzheimers Dement.* 13, p1268–1269. doi: 10.1016/j.jalz.2017.06.1897

Li, S., Okonkwo, O., Albert, M., and Wang, M. C. (2013). Variation in variables that predict progression from MCI to AD dementia over duration of follow-up. *Am. J. Alzheimers Dis.* 2, 12–28. doi: 10.7726/ajad.2013.1002

Liu, M., Zhang, D., Yap, P. T., and Shen, D. (2012). Tree-guided sparse coding for brain disease classification. *Med. Image Comput. Comput. Assist. Interv.* 15(Pt 3), 239–247. doi: 10.1007/978-3-642-33454-2_30

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., and Sarazin, M. (2010). Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. *Neuroradiology* 51, 73–83. doi: 10.1007/s00234-008-0463-x

Martinez-Murcia, F. J., Ortiz, A., Gorriz, J. M., Ramirez, J., and Castillo-Barnes, D. (2019). Studying the manifold structure of Alzheimer's Disease: a deep learning approach using convolutional autoencoders. *IEEE J. Biomed. Health Inform.* 1-1. doi: 10.1109/JBHI.2019.2914970

Miksad, R. A., and Abernethy, A. P. (2018). Harnessing the Power of Real-World Evidence (RWE): a checklist to ensure regulatory-grade data quality. *Clin. Pharmacol. Ther.* 103, 202–205. doi: 10.1002/cpt.946

Moeller, J. (2015). A word on standardization in longitudinal studies: don't. *Front. Psychol.* 6:1389. doi: 10.3389/fpsyg.2015.01389

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., and Verbeke, G. (2014). *Handbook of Missing Data Methodology*. New York, NY: Chapman and Hall/CRC. doi: 10.1201/b17622

Moreno-Betancur, M., Leacy, F. P., Tompsett, D., and White, I. (2017). *mice: The NARFCS Procedure for Sensitivity Analyses*.

Moulder, K. L., Snider, B. J., Mills, S. L., Buckles, V. D., Santacruz, A. M., Bateman, R. J., et al. (2013). Dominantly inherited Alzheimer network: facilitating research and clinical trials. *Alzheimers Res. Ther.* 5:48. doi: 10.1186/alzrt213

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., et al. (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N Am.* 15, 869–877. doi: 10.1016/j.nic.2005.09.008

Nettiksimmons, J., DeCarli, C., Landau, S., and Beckett, L. (2014). Biological heterogeneity in ADNI amnestic mild cognitive impairment. *Alzheimers Dement.* 10, 511–521.e1. doi: 10.1016/j.jalz.2013.09.003

Neu, S. C., Crawford, K. L., and Toga, A. W. (2012). Practical management of heterogeneous neuroimaging metadata by global neuroimaging data repositories. *Front. Neuroinform.* 6:8. doi: 10.3389/fninf.2012.00008

Neville, J., Kopko, S., Romero, K., Corrigan, B., Stafford, B., LeRoy, E., et al. (2017). Accelerating drug development for Alzheimer's disease through the use of data standards. *Alzheimer's Dement.* 3, 273–283. doi: 10.1016/j.trci.2017.03.006

Ning, J., Qin, J., and Shen, Y. (2010). Nonparametric tests for right-censored data with biased sampling. *J. R. Stat. Soc. Series B Stat. Methodol.* 72, 609–630. doi: 10.1111/j.1467-9868.2010.00742.x

O'Bryant, S. E., Gupta, V., Henriksen, K., Edwards, M., Jeromin, A., Lista, S., et al. (2015). Guidelines for the standardization of preanalytic variables for blood-based biomarker studies in Alzheimer's disease research. *Alzheimers Dement.* 11, 549–560. doi: 10.1016/j.jalz.2014.08.099

Oxtoby, N. P., and Alexander, D. C. (2017). Imaging plus X: multimodal models of neurodegenerative disease. *Curr. Opin. Neurol.* 30, 371–379. doi: 10.1097/WCO.0000000000000460

Oxtoby, N. P., Young, A. L., Cash DM Benzinger, T. L. S., Fagan, A. M., Morris, J. C., et al. (2018). Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain* 141, 1529–1544. doi: 10.1093/brain/awy050

Peng, D., Shi, Z., Xu, J., Shen, L., Xiao, S., Zhang, N., et al. (2016). Demographic and clinical characteristics related to cognitive decline in Alzheimer's disease in China: a multicenter survey from 2011 to 2014. *Medicine* 95:26. doi: 10.1097/MD.0000000000003727

Petrella, J. R., Hao, W., Rao, A., and Doraiswamy, P. M. (2019). Computational causal modeling of the dynamic biomarker cascade in Alzheimer's disease. 2019:6216530 *Comput. Math. Methods Med.* doi: 10.1155/2019/6216530

Pourhoseingholi, M. A., Baghestani, A. R., and Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterol. Hepatol. Bed Bench* 5, 79–83.

Prabhakaran, G., and Bakshi, R. (2018). Analysis of structure and cost in an American longitudinal study of Alzheimer's disease. *J. Alzheimers Dis. Parkinsonism* 8:411. doi: 10.4172/2161-0460.1000411

Prinja, S., Gupta, N., and Verma, R. (2010). Censoring in clinical trials: review of survival analysis techniques. *Indian J. Community Med.* 35, 217–221. doi: 10.4103/0970-0218.66859

Quartagno, M., and Carpenter, J. (2016). *jomo: A Package for Multilevel Joint Modelling Multiple Imputation.* R package version 2.

Rao, A., Lee, Y., Gass, A., and Monsch, A. (2011). Classification of Alzheimer's disease from structural MRI using sparse logistic regression with optional

spatial regularization. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 4, 499–502. doi: 10.1109/IEMBS.2011.6091115

Reitz, C. (2016). Toward precision medicine in Alzheimer's disease. *Ann. Transl. Med.* 4:107. doi: 10.21037/atm.2016.03.05

Ricciarelli, R., and Fedele, E. (2017). The amyloid cascade hypothesis in Alzheimer's disease: it's time to change our mind. *Curr. Neuropharmacol.* 2017, 926–935. doi: 10.2174/1570159X15666170116143743

Rodriguez-Esteban, R. (2015). Biocuration with insufficient resources and fixed timelines. *Database* 2015:bav116. doi: 10.1093/database/bav116

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. doi: 10.1038/s42256-019-0048-x

Samtani, M. N., Raghavan, N., Shi, Y., Novak, G., Farnum, M., and Lobanov, V. (2013). Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. *Br. J. Clin. Pharmacol.* 75, 146–161. doi: 10.1111/j.1365-2125.2012.04308.x

Schott, J. M., and Petersen, R. C. (2015). New criteria for Alzheimer's disease: which, when and why? *Brain* 138(Pt 5), 1134–1137. doi: 10.1093/brain/awv055

Simon, C., Davidsen, K., Hansen, C., Seymour, E., Barnkob, M. B., and Olsen, L. R. (2018). BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics* 19:57. doi: 10.1186/s12859-019-2607-x

Singh, M., Murthy, A., and Singh, S. (2015). Prioritization of free-text clinical documents: a novel use of a bayesian classifier. *JMIR Med. Inform.* 3:e17. doi: 10.2196/medinform.3793

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., et al. (2011a). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 280–292. doi: 10.1016/j.jalz.2011.03.003

Sperling, R. A., Jack, C. R. Jr., and Aisen, P. S. (2011b). Testing the right target and right drug at the right stage. *Sci. Transl. Med.* 3:111cm33. doi: 10.1126/scitranslmed.3002609

Thanh, N.,Tran, T., Drab, K., and Daszykowski, M. (2013). Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *J. Chemom. Intell. Lab. Syst.* 120, 92–96. doi: 10.1016/j.chemolab.2012.11.006

Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *J. Royal Stat. Soc.* 63, 411–423. doi: 10.1111/1467-9868.00293

Tombaugh, T. N. (2005). Test-retest reliable coefficients and 5-year change scores for the MMSE and 3MS. *Arch. Clin. Neuropsychol.* 20, 485–503. doi: 10.1016/j.acn.2004.11.004

Toschi, N., Lista, S., Baldacci, F., Cavedo, E., Zetterberg, H., Blennow, K., et al. (2019). Biomarker-guided clustering of Alzheimer's disease clinical syndromes. *Neurobiol. Aging* 83, 42–53. doi: 10.1016/j.neurobiolaging.2019.08.032

Vermunt, L., Veal, C. D., Ter Meulen, L., Chrysostomou, C., van der Flier, W., Frisoni, G. B., et al. (2018). European prevention of Alzheimer's dementia registry: recruitment and pre screening approach for a longitudinal cohort and prevention trials. *Alzheimers. Dement.* 14, 837–842. doi: 10.1016/j.jalz.2018.02.010

Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P. M., and Ye, J. (2013). "Multi-source learning with block-wise missing data for Alzheimer's disease prediction,". in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY: ACM), 185–193. doi: 10.1145/2487575.2487594

Younesi, E., and Hofmann-Apitius, M. (2013). From integrative disease modeling to predictive, preventive, personalized and participatory (P4) medicine. *EPMA J.* 4:23. doi: 10.1186/1878-5085-4-23

Young, A. L., Marinescu, R. V., Oxtoby, N. P., Bocchetta, M., Yong, K., Firth, N. C., et al. (2018). Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat. Commun.* 9:4273. doi: 10.1038/s41467-018-05892-0

Young, A. L., Oxtoby, N. P., Huang, J., Marinescu, R. V., Daga, P., Cash, D. M., et al. (2015). Multiple orderings of events in disease progression. *Inf. Process. Med. Imaging* 24, 711–722. doi: 10.1007/978-3-319-19992-4_56

Young, J., Modat, M., Cardoso, M. J., Mendelson, A., Cash, D., and Ourselin, S. (2013). Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin.* 2, 735–745. doi: 10.1016/j.nicl.2013.05.004

Zhang, D., Wang, Y., Zhou, L., Yuan, H., and Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008

Zhang, J., Zhou, W., Cassidy, R. M., Su, H., Su, Y., Zhang, X. (2018). Risk factors for amyloid positivity in older people reporting significant memory concern. *Comprehensive Psychiatry* 80, 126–131. doi: 10.1016/j.comppsych.2017.09.015

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Golriz Khatami, Robinson, Birkenbihl, Domingo-Fernández, Hoyt and Hofmann-Apitius. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*