Data-driven modelling of neurodegenerative disease progression: thinking outside the black box

Alexandra L Young[1,2]*, Neil P Oxtoby[1]*, Sara Garbarino[3], Nick C Fox[4], Frederik Barkhof[1,5], Jonathan M Schott[4], Daniel C Alexander[1]

*Joint first authors

[1]UCL Centre for Medical Image Computing, Department of Computer Science, University College London, UK
[2]Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK
[3]Life Science Computational laboratory, IRCCS Ospedale Policlinico San Martino, Genova, Italy
[4]Dementia Research Centre, UCL Queen Square Institute of Neurology, University College London, UK
[5]Department of Radiology & Nuclear Medicine, Amsterdam University Medical Center, Netherlands

ORCID
ALY: https://orcid.org/0000-0002-7772-781X
NPO: https://orcid.org/0000-0003-0203-3909
SG: https://orcid.org/0000-0002-3583-3630
NCF: https://orcid.org/0000-0002-6660-657X
FB: https://orcid.org/0000-0003-3543-3706
JMS: https://orcid.org/0000-0003-2059-024X
DCA: https://orcid.org/0000-0003-2439-350X

Abstract

Data-driven disease progression models are an emerging set of computational tools that reconstruct disease timelines in long-term chronic diseases, providing unique insights into disease processes and their underlying mechanisms. Such methods combine *a priori* human knowledge/assumptions with large-scale data processing and parameter estimation to infer long-term disease trajectories from short-term data. In contrast to "black box" machine learning tools, disease progression models typically require less data and are inherently interpretable, thereby aiding disease understanding in addition to enabling classification, prediction, and stratification. While initially developed by the statistics and machine learning communities for neurodegenerative disease applications, the techniques have reached a technological maturity level allowing adoption by the wider scientific community for use across a range of neuroscience and non-neuroscience applications. Here we place the current landscape of data-driven disease progression models in a general framework and discuss the enhanced utility of constructing a disease timeline compared to wider machine-learning tools that construct static disease profiles. We review the insights they have enabled across multiple neurodegenerative diseases for applications such as determining temporal trajectories of disease biomarkers, testing hypotheses about disease mechanisms, and uncovering disease subtypes. Finally, we outline key areas for technological development and discuss potential pathways and barriers to integrating disease progression models in clinical practice and trial settings.

Introduction

Neurodegenerative diseases characteristically evolve over timescales of many years with long preclinical/prodromal periods. Example neurodegenerative diseases with a long preclinical phase include Alzheimer's disease[1], Parkinson's disease[2], frontotemporal dementia[3], and Huntington's disease[4]. Neurodegenerative diseases are commonly associated with a stereotypical temporal pattern of disease biomarker changes – a ***disease timeline***. This timeline reflects underlying measureable disease processes (biomarkers) that define the condition, often being unique to a particular condition or subtype[5]. Disease timelines provide a mechanism for disease classification across different patients and disease stages, a window into disease biology, and a framework for cohort selection in clinical trials and research. More broadly, many neurological and non-neurological conditions evolve over long timescales, often including a preclinical stage, and are associated with a stereotypical timeline of disease biomarker changes. Examples include multiple sclerosis[6] and lung diseases such as Chronic Obstructive Pulmonary Disease[7].

Mapping the timeline of a neurodegenerative disease would be relatively simple if it were possible to obtain densely sampled start-to-end longitudinal measurements from a set of individuals known to be on a similar disease trajectory. The reality of chronic illness does not afford such luxury. First, the requirement for dense decades-long assessment is impractical at scale, even when ignoring preclinical stages. Examinations are usually inconvenient (e.g. travelling to a memory clinic or imaging centre), often invasive (e.g. lumbar puncture), and expensive. Moreover, even when following a few individuals is possible, earlier technologies become outdated and data-consistency is low. Second, many neurodegenerative diseases are sporadic and thus cases are difficult to identify prior to symptoms, which may arise many years after the pathological process has begun. Measurements thus often undersample the important pre-symptomatic phase. Finally, the heterogeneity of typical patient populations and complexity of disease mechanisms exacerbate these problems. Studies require large cohorts to capture the variability of patient trajectories, and diverse sets of biomarkers to inform on the variety of disease processes involved. Real-world patient datasets thus comprise predominantly cross-sectional and short-term longitudinal biomarker measurements, are sparser towards the beginning of the disease, and are heterogeneous in nature.

Data-driven disease progression models are an emerging set of computational tools that infer long-term disease timelines from short-term biomarker data. In contrast to classification or clustering tools, which construct a static average profile for a (sub)group, data-driven disease progression models infer a timeline describing the sequential progression or evolution of a disease over time. The timelines inferred by data-driven disease progression models have a range of applications and are increasingly being used in research studies to provide biological insights into neurodegenerative diseases and as a stratification tool. While initially developed for neurodegenerative disease applications, data-driven disease progression models have broader applicability across a range of long-term chronic conditions and are starting to be adopted for use across a wider range of neuroscience and non-neuroscience applications.

In this review, we summarise the current state-of-the-art in data-driven disease progression modelling. Our review is structured as follows. We first define the term 'data-driven disease progression model' and provide a framework for data-driven disease progression modelling, broadly segregating disease progression models into phenomenological and pathophysiological models: the former aim to capture common trajectories of disease biomarkers without considering underlying mechanisms, while the latter aim to explain disease timelines in terms of biological and physical processes and characteristics. We describe the broader landscape of approaches that motivated and inspired the development

of data-driven disease progression models. We review the state of the art for phenomenological and pathophysiological models and summarise the biological insights provided to date across multiple neurodegenerative diseases. Finally, we discuss further opportunities for technical development and how to realise the full potential of data-driven disease progression models as tools for disease understanding and management.

Data-driven disease progression models

Definition

Data-driven disease progression models are a family of statistical and machine learning tools developed to learn long-term disease biomarker timelines of chronic diseases from short-term data without requiring prior knowledge of an individual's disease stage. This enables biomarker changes to be mapped at a fine-grained temporal resolution.

The term 'data-driven disease progression model' is inconsistently used in the literature to refer to a range of models. For the purposes of our review, we define data-driven disease progression models as having two key features:

1. **They construct a *data-driven disease timeline*.**
   A data-driven disease timeline is a generative model of how a disease evolves over time that is indexed by a ***data-driven disease time axis*** – an inferred model-based time axis that measures an individual's position along an expected average disease timeline. This feature enables the inference of timelines with a fine-grained temporal resolution.
2. **They are directly informed by measured data.**
   The use of *in vivo* biomarkers ensures quantitative disease timelines that provide biological insight, enable patient staging and stratification, and/or provide predictions of long-term disease progression.

Current data-driven disease progression models segregate into two broad categories: **phenomenological** and **pathophysiological** models. Phenomenological models aim to capture common trajectories of disease biomarkers without considering underlying mechanisms. Pathophysiological models aim to explain disease timelines in terms of biological and physical processes and characteristics, such as what determines vulnerability to disease and the spread of pathology.

Framework

Figure 1 outlines a framework for data-driven disease progression modelling. Data-driven disease progression models use a generative disease progression model and a set of constraints informed by human insight to infer a data-driven disease time axis and the shape of biomarker trajectories along it. The data-driven time axis is a model-based time axis that describes an expected average disease timeline, temporally realigning individuals relative to this timeline. This enables short-term data to inform long-term disease trajectories, whilst a set of trajectory constraints informed by human insights enable reconstruction from noisy medical datasets and ensure interpretability of outputs. In contrast to classical regression techniques (see **Broader Landscape and Historical Context: Regression** below), which are inherently limited by the temporal resolution of the disease staging measure used to

position an individual along the time axis, data-driven disease progression models can reconstruct disease timelines at a more fine-grained temporal resolution.

The framework describes both phenomenological and pathophysiological models. Phenomenological models typically use weaker constraints relating only to the expected shape of the trajectories. This requires complex fitting strategies that temporally realign short-term snapshots to infer the parameters of the long-term biomarker trajectories, as Figure 1 illustrates. In contrast, many pathophysiological models are so highly constrained that they can use a simple fitting approach that compares an expected trajectory for a set of parameters to an average end-stage disease pattern. Those models can be described in the framework as aligning the trajectories to a single late stage timepoint. Recent pathophysiological models (see last paragraph of **Pathophysiological models: Dynamical systems models** below) use the strategy of phenomenological models of realigning short-term snapshots along the full disease time course to enable the fitting of more complex (less constrained) models. Our framework places the broad spectrum of phenomenological and pathophysiological models in a common paradigm.

Broader Landscape and Historical Context

Inferring neurodegenerative disease timelines from real-world datasets, which are predominantly cross-sectional and short-term longitudinal, presents a technical challenge. Here we describe the broader landscape and historical context that inspired the development of data-driven disease progression modelling for reconstructing long-term progression from short-term data.

Neuropathological staging systems. Neuropathological staging systems[8–14] aim to reconstruct long-term disease progression patterns from cross-sectional data but are not directly applicable to *in vivo* biomarker data. The Braak and Braak staging system[8] hypothesises that Alzheimer's disease spreads over the brain from region to region in a stereotypical manner and uses this assumption to derive a neuropathological staging system from cross-sectional data based on the perceived frequency with which different regions are affected across patients. Specifically, if we observe that pathology B is often present without pathology A, but only rarely observe A without B, we can infer that the disease usually produces B before A. This idea underpins the mathematical construction of the event-based model (see **Phenomenological models: Discrete models**), enabling reconstruction of longitudinal progression from entirely cross-sectional data. More generally, the observation of sequential progression of neurodegenerative diseases from region to region and the idea of reconstructing longitudinal progression from cross-sectional (or short-term longitudinal) data inspired a broad range of disease progression models and supported the assumption made by most disease progression models of monotonic progression across populations.

Hypothetical biomarker models. Hypothetical models[15–17] describe expected timelines of Alzheimer's disease biomarker progression but are not quantitative. For example, Jack et al.[15,16] describe a hypothesised temporal pattern of *in vivo* biomarker evolution in Alzheimer's disease based on their review of available data, proposing a model in which amyloid biomarkers become abnormal first, followed by neurodegenerative biomarkers and cognitive symptoms, with neurodegenerative biomarkers correlating with clinical symptom severity. Hypothetical models have been highly influential in framing the debate surrounding the expected pattern of Alzheimer's biomarker changes and etiopathological mechanism. The unmet need for quantitative versions of these models, based on observed data, motivated

the development of data-driven disease progression models, and inspired some of their design choices. Hypothetical models remain an important method of integrating and debating overall research findings and are continually being proposed and updated in response to the literature[18].

Regression. A simple statistical approach to inferring long-term progression from short-term data is to use regression to find an average trajectory across individuals, i.e., to chart an individual's biomarker measurements against their position along the disease time course. Classical regression techniques rely on a directly measured empirical disease time axis requiring knowledge of an individual's stage along the disease time axis. This stage is not well defined for long-term chronic diseases — especially during the pre-symptomatic phase, limiting the temporal resolution of the inferred disease timelines. We highlight that, although they are data-driven, classical regression techniques do not meet our definition of a data-driven disease progression model because they rely on an empirical disease time axis. This limits their temporal resolution to the resolution of the disease staging measure used to index the disease time axis. Example disease-staging measures that have been used include:

- Chronological age. Biomarker trajectories have been charted relative to chronological age[19]; however, age at onset of neurodegenerative diseases is variable.
- Clinical staging. Clinical (symptomatic) staging measures have been used to approximate disease progression (e.g. [20–25]) but are typically crude (e.g. 'mild', 'moderate', 'severe') and their reliance on cognitive test scores prevents monitoring of progression during pre-symptomatic disease stages.
- Biomarker indexing. Many studies have mapped biomarker patterns relative to a single biomarker as a proxy of disease stage[26–29], such as charting cortical and hippocampal atrophy against a cognitive test score[28]. However, most biomarkers are sensitive only to a particular disease stage and may be non-specific to the disease of interest.
- Expected age-of-onset in dominantly inherited conditions. More fine-grained temporal resolution has been mapped for dominantly inherited genetic diseases[1,3] where parental age-of-onset and/or genetic markers can provide an approximate disease-stage benchmark. However, such staging systems are imperfect[30] and do not generalise to sporadic disease.
- Time-to-conversion. More recently, several observational cohort studies have reached a duration long enough to observe conversion between diagnoses in moderately sized populations, enabling staging of individuals by time to diagnosis[31–34]. However, such studies are limited by the accuracy of time of diagnosis and under sample the pre-symptomatic disease phase.

Imperfect disease staging systems further limit the ability to characterise the heterogeneity of neurodegenerative diseases. Clustering is frequently used for this purpose[35,36] but requires an accurate measure of disease stage to avoid conflating disease subtypes and stages.

Phenomenological Models

Phenomenological models (**Figure 2**) jointly learn disease biomarker trajectories and a data-driven disease time axis. These disease signatures have direct application for staging and stratification (see **Applications**) through comparison of an individual's biomarker values to model(s) to find the best fitting stage and/or stratum. Phenomenological models also support some basic mechanistic insights. For example, if a model indicates that biomarker A

becomes abnormal before biomarker B, one might infer that underlying biological processes follow a similar temporal relationship.

We broadly divide phenomenological models into four classes: discrete, continuous, spatiotemporal, and subtyping models. **Figure 2** provides examples. **Table 1** summarises the attributes of each class of phenomenological model and compares them to pathophysiological models. Phenomenological models are generally designed to take tabular data as input, i.e. spreadsheets of per-patient biomarker data sets derived from clinical studies, which may be cross-sectional or short-term longitudinal. More complex models, such as spatiotemporal models, sometimes work directly with raw data (e.g. image data sets) rather than derived biomarkers. Model output typically consists of a set of biomarker trajectories along a data-driven disease time axis. Human knowledge constrains the shape of those trajectories, e.g., by specifying a parametric functional form or a monotonicity constraint. Such constraints are relatively weak compared to the much stronger (*a priori*) constraints used by pathophysiological models.

Discrete Models

Discrete models describe disease progression as transitions through a series of states. Each state consists of an expected set of biomarker values, or ranges/distributions of values. Simple models (see **The event-based model**) impose a strict ordering of states, which defines the disease time axis as a sequence of disease stages with every subject passing through every stage. More complex models (see **Extensions to the event-based model** and **Hidden Markov Models**) may use a partial ordering of states so that different individuals follow different pathways and don't necessarily transition through every state.

The event-based model. The event-based model[37] (**Figure 2A**) is one of the earliest disease progression models. Mathematically, the event-based model describes disease progression as an event sequence – a series of irreversible events representing transitions of a series of biomarkers from 'normal' to 'abnormal'. Thus, at stage N the first N biomarkers are abnormal, while all other biomarkers remain normal. Estimation seeks the probabilistic ordering of such events, with a probabilistic model for each biomarker transition. The general principle of the event-based model is similar to that used by (heuristic) neuropathological staging systems: if we observe abnormality in biomarker A without abnormality in biomarker B more often than vice versa, we infer that A typically becomes abnormal before B. However, the event-based model formalises this idea within a probabilistic framework, acknowledging that there will be a distribution of 'normal' and 'abnormal' values across the population and handling uncertainty. The event-based model formalises the uncertainty in the population-level event sequence and uncertainty in the stage of an individual along the event sequence.

Extensions to the event-based model. A range of enhancements have been developed enabling the use of the event-based model in a broader range of applications. Initially, the event-based model was demonstrated in genetically inherited diseases, requiring a well-defined control population. Young et al.[38] developed a mixture modelling approach enabling broader application to sporadic disease datasets. The original event-based model assumed parametric distributions. Firth et al.[39] proposed a kernel density mixture modelling approach, facilitating application to non-parametric data, such as cognitive test scores. The event-based model describes events as the transition from a 'normal' to an 'abnormal' level. Young et al. introduced z-score[40] and ordinal[41] versions, modelling the transitions of biomarkers between different scores for continuous (e.g. regional brain volumes) and ordinal (e.g. neuropathological or clinical ratings) data respectively. Huang et al.[42] and Venkatraghavan et al.[43] each developed versions of the event-based model that use a generalised Mallows model to describe a distribution of event orderings, capturing greater variability in the population-level event sequence than the original event-based model. The event-based

model assumes that each event in the sequence occurs at a distinct stage, recent extensions enable modelling of sets of events that occur simultaneously[44,45]. The event-based model describes an ordinal disease time axis comprising discrete stages with arbitrary duration. Du et al.[46] reformulate the event-based model to have a continuous time axis with an arbitrary timescale, similar to pseudotime approaches (see **Continuous Models: Pseudotime approaches** below). Wijeratne et al.[47,48] developed an event-based modelling approach that incorporates short-term longitudinal data to enable estimation of the absolute timescale of the time axis (see **Hidden Markov Models** section below).

Hidden Markov Models. Hidden Markov Models are a classical machine learning approach that has been adapted for the task of disease progression modelling. Hidden Markov Models describe the temporal evolution of a set of states, modelling the expected series of future states under the assumption that each state depends only on the previous state. The use of Hidden Markov Models for disease progression modelling is challenging as they learn many parameters and have no natural time directionality (individuals can move from/to any state from any other). Hidden Markov Models have been adapted for disease progression modelling by using short-term longitudinal data to impose time directionality and restricting the number of states to ten or fewer[49,50]. The event-based model can be thought of as a Hidden Markov Model with a monotonicity constraint that enables the inference of the states (stages) from purely cross-sectional data. The recently introduced temporal event-based model[47,48] incorporates ideas from Hidden Markov Modelling into the event-based model to infer the time between events using short-term longitudinal data.

Continuous Models

Continuous models describe biomarker trajectories as a continuous function of a data-driven disease time axis. Continuous models can be estimated from cross-sectional data to derive timelines with an arbitrary timescale (see **Pseudotime approaches** below). However, in practice most use short-term longitudinal data to reconstruct timelines with an absolute timescale (see **Differential equation models** and **Latent-time regression approaches** below). The current literature includes three key approaches: differential equation models that typically estimate biomarker trajectories independently; and pseudotime approaches and latent-time regression models that jointly estimate a set of biomarker trajectories and a common time axis.

Differential equation models. Differential equation models model short-term subject-level changes as estimates of the derivative of a long-term group-level biomarker trajectory, which is inferred by integrating/inverting the model. Mathematically, this takes on the structure of a phase plane in physics, i.e., a model of (biomarker) velocity versus position. Velocity and position are calculated per individual datum (e.g., patient observation), with the group-level model fit inverted into a trajectory. Several differential equation models have been proposed between 2011 and 2014[28,51–54] and were influential in characterising the overall timescale of Alzheimer's trajectories (see **Applications**), particularly the pre-symptomatic phase. However, most phenomenological differential equation models describe the trajectories of individual biomarkers separately. Thus, unlike most classes of disease progression model, they do not estimate a disease timeline common to all biomarkers and do not provide a single disease stage for individuals out-of-the-box.

Pseudotime approaches. Pseudotime approaches[55] (also referred to as trajectory inference methods) are a set of computational techniques developed by the single-cell transcriptomics community to study a variety of cellular dynamic processes (e.g. the cell cycle, cell differentiation, cell activation). Pseudotime approaches order cells along a trajectory based on similarities in their expression patterns, indexing cells by a disease 'pseudotime', which measures their relative position along the trajectory on an arbitrary timescale. Pseudotime approaches have recently been adapted for inferring continuous

data-driven disease timelines from purely cross-sectional data. Saint-Jalmes et al.[56] tested the performance of a simple pseudotime approach involving principal components analysis that showed promise for recapitulating some basic features of more complex models. Iturria-Medina et al.[57–59] developed a contrastive trajectory inference approach for estimating disease subtypes and timelines, inspired by pseudotime approaches (see **Phenomenological models: Disease subtype models**).

Latent-time regression approaches. Latent-time regression approaches (**Figure 2B**) jointly estimate a data-driven disease time axis and a set of biomarker trajectories indexed along this common time axis. Mathematically, they take on the structure of a multivariate mixed-effects model with parametric or nonparametric dependence upon a data-driven disease axis. Joint inference of the data-driven disease time axis and biomarker trajectories is performed by iteratively fitting a set of biomarker trajectories and temporally realigning individuals to improve their alignment with the fitted biomarker trajectories. This process of temporally realigning individuals is often referred to as a 'time-shift'. The Disease Progression Score model[60] includes group-level sigmoidal biomarker trajectories as a function of individual-level disease progression scores that transform chronological age by an individual's rate of progression and age of disease onset. Donohue et al.[61] model disease trajectories as continuously differentiable monotonic functions, allowing for a Gaussian distributed time-shift per individual, and modelling an individualised intercept and slope (random effects) for each biomarker. Li et al.[62] extended this work to accommodate fixed effects and use a Bayesian framework to model uncertainty. Lorenzi et al.[63] introduced a nonparametric Gaussian process model allowing variable uncertainty along the trajectory. Raket et al.[64] developed a parameterised mixed-effects model tailored specifically for cognitive data. A range of latent-time regression approaches have been proposed to learn spatiotemporal image trajectories (see **Phenomenological Models: Spatiotemporal Models** below), many of which naturally handle scalar biomarker data, e.g.[65–71].

Spatiotemporal Models

Spatiotemporal models[65–70,72–76] (**Figure 2C**) are typically based on ideas similar to latent-time regression approaches but operate in a high dimensional space to enable the modelling of either full images, shape changes of specific brain regions (such as the shape of the hippocampus), or image feature maps (such as maps of cortical thickness on the cortical surface). These models emerged from the medical image registration community where warping images to a common space is an early step in group analyses[77]. Multi-modality models typically handle combinations of shape changes, feature maps and scalar values. Voxel-wise models handle full images but are typically limited to a single modality.

Multi-modality models. Schiratti et al.[65–67] were the first to develop a spatiotemporal model with a time-shift, inspired by earlier work (e.g. [72,73]) introducing time reparameterisations in other spatiotemporal modelling contexts. Schiratti et al. proposed a range of spatiotemporal models that were later formalised into a single framework[68] estimating parameters for the time-shift of an individual, the rate of progression of an individual, and the relative positioning of each biomarker along the data-driven disease time axis. Whilst theoretically applicable to complex image data, the demonstrations of these models were restricted to scalar biomarkers. Koval et al.[70] adapted this framework to develop the 'Disease Course Mapping' approach, which has been demonstrated on combinations of scalar biomarkers, vertex-wise data, and shape data (see example in **Figure 2C**). Louis et al.[69] proposed a method that uses a deep generative network to learn trajectories in a low-dimensional space and then map them to the observation space. Abi Nader et al.[76] developed a spatiotemporal version of the Gaussian process model of Lorenzi et al.[63]

Voxel-wise models. Bilgel et al.[74] and Marinescu et al.[75] each developed voxel-wise latent-time regression approaches that reconstruct high resolution disease progression patterns within a single modality. The approach of Bilgel et al.[74] accounts for spatial correlation between neighbouring voxels, producing a smooth spatial map of disease progression. In contrast, the approach of Marinescu et al.[75] uses weaker spatial constraints, clustering together vertices with common progression dynamics regardless of their spatial proximity. This approach offers the potential to identify smaller localised changes but may be more sensitive to noise in the images.

Disease subtype models

These models (**Figure 2D**) relax the assumption made by earlier phenomenological models of a single common disease timeline by combining ideas from clustering and disease progression modelling to estimate distinct data-driven disease timelines for multiple subgroups.

Discrete disease subtyping models. The Subtype and Stage Inference (SuStaIn)[40] algorithm aims to estimate both disease subtype and timeline simultaneously. The original SuStaIn implementation combines a discrete disease progression model (z-score event-based model) with clustering to estimate simultaneously a set of disease subtypes, the set of biomarker trajectories defining each subtype, and a data-driven (ordinal) disease time axis. This information can subsequently assign individuals to a subtype and stage. More recent developments and applications demonstrate flexibility of the SuStaIn algorithm to use alternative disease progression models[40,41,78] and to incorporate longitudinal data to estimate the absolute timescale of each subtype timeline[79].

Continuous disease subtyping models. Contrastive trajectory inference[57–59] is a continuous disease subtyping model for cross-sectional data based on pseudotime approaches that learns disease subtypes with distinct timelines on an arbitrary timescale. Contrastive trajectory inference consists of three steps: feature selection, dimensionality reduction using contrastive principal components analysis, and pseudotemporal ordering to obtain individual disease scores. This pseudotime describes an individual's distance from a control population in the contrasted principal components space, with a minimal spanning tree being computed to group together subgroups of individuals that share a common trajectory. Similarly, filtered trajectory recovery[46] is a continuous extension to the event-based model and SuStaIn that enables recovery of disease subtypes with distinct timelines with an arbitrary timescale. Alternative continuous disease subtyping models[80,81] use short-term longitudinal data to estimate an absolute timescale for each subtype timeline. SubLign[80] is a two-stage algorithm that first learns individual-level time series by using a deep generative model to disentangle temporal variation in the data due to disease severity at baseline from spatial variation due to subtype, and then clusters these individual-level time series using k-means clustering to obtain a subtype identity for each individual. Poulet and Durrleman[81] developed a clustering extension to Disease Course Mapping[70] based on mixture modelling, jointly estimating subtypes and a non-linear mixed effects model for each subtype.

Pathophysiological Models

Pathophysiological models (**Figure 3** and **Figure 4**) describe expected disease timelines in terms of underlying pathophysiological processes and patient characteristics. The models estimate the characteristics of assumed pathophysiological processes that generate disease timelines that best predict observed biomarker measurements, typically focussing more on biological hypothesis-testing and insight than on disease phenomenology.

Pathophysiological models can provide evidence for/against competing hypotheses about disease mechanisms.

This section first discusses the concepts of pathogen appearance and spreading, which are central to all classes of pathophysiological model. We then review the state of the art in pathophysiological modelling, broadly dividing pathophysiological models into network models, dynamical systems models, and models of mechanistic combinations (which can include network models and dynamical systems models).

Table 1 summarises the attributes of each class of pathophysiological models, contrasting them with phenomenological models. Pathophysiological models typically take as input a model or approximation of a topological property of the brain, such as a connectivity map[82] or gene-expression map e.g.[83] (often based on an average template from healthy controls), which is assumed to drive or mediate pathophysiology according to human knowledge or hypotheses. Leveraging relatively strong *a priori* constraints, a model outputs a set of expected biomarker trajectories along a data-driven disease timeline, together with estimates of any pathophysiological parameters tuned by the model. Network models have very high constraints, with the expected pathology pattern being entirely defined by connectivity metrics. The constraints in a dynamical systems model are slightly weaker but still high, typically estimating only a few key parameters of a pathophysiological process. Models of mechanistic combinations aim to describe the relative effects, and possibly interactions, of multiple mechanisms simultaneously. Their constraint level varies according to the number of mechanisms and pathophysiological parameters per mechanism; however, we describe a typical model as being moderately constrained.

Appearance and spreading

Pathophysiological models are often used to model the appearance and/or spread of pathology in and between brain regions. **Figure 3** illustrates this conceptually, showing how characteristics of the brain connectivity network can be linked to the appearance and spread of pathology. Appearance mechanisms (**Figure 3,** upper rows) include “hub vulnerability” (**Figure 3**, first row), where the stress of high usage increases disease susceptibility[84–86]; and “hub sparing” (**Figure 3**, second row), where low usage/connectivity increases disease susceptibility[87,88]. Spreading mechanisms (**Figure 3** lower rows) include prion-like pathology diffusion from an epicentre (**Figure 3** third row), which can be constrained along white-matter fibres[89–91] and influenced by the orientation of white matter fibres[92], or unconstrained[93] (**Figure 3** fourth row).

Network models

Network models implement mechanistic hypotheses by associating brain network properties with disease progression timelines. Models typically use functional or structural connectivity matrices from (often healthy) functional MRI or diffusion MRI tractography[82,94–101], based on the observation that the pathways of protein propagation largely overlap with functional or structural brain networks[13,88,102], but may also use other estimates of brain connectivity networks[103,104]. The seminal study of Zhou et al.[82] aimed to test various mechanistic hypotheses of neurodegenerative pathology – transneuronal or “prion-like” spread[102,105,106], nodal stress[85,86,107], trophic failure[108], and shared vulnerability[87,109–112] – via graph theory metrics of the brain’s functional connectivity. Mathematically, hub vulnerability was modelled using network centrality/segregation metrics, which are high/low in brain hubs and low/high in isolated regions, respectively. Pathology appearance and spreading was modelled using graph distance from an epicentre. Each hypothesised mechanism produces a competing expected disease timeline, the evidence for which is evaluated by comparison with late-

stage atrophy severity patterns in Alzheimer's disease. The authors found that network distance had the strongest evidence, supporting prion-like spreading from an epicentre.

Subsequent studies based on functional connectivity reported similar findings that favour the "prion-like spreading hypothesis" over other network-based mechanistic hypotheses[96–98]. Others use tractography-based structural connectivity[94,95], which estimates physical connections (rather than functional correlations) between brain regions along which pathogens could spread but draw similar conclusions on the prion hypothesis in Alzheimer's disease. Indeed, it has been observed that the strength of resting-state functional connectivity closely correlates with structural connectivity strength[113], indicating that functional networks are neuronal in origin. These studies mostly use the same set of network metrics originally proposed by Zhou[82], except for one[96] who used a participation coefficient between subnetworks, which might correlate with metabolic activity[114].

Dynamical systems models

Dynamical systems models aim to emulate biophysical spatiotemporal processes of protein appearance, spreading, and clearance in neurodegenerative diseases. These models typically encode protein dynamics in the kinetic parameters of a system of differential equations[92,115–124]. Mathematically, dynamical systems models estimate a timeline that is completely driven by the dynamics inferred from an underlying set of differential equations, and thus mostly differ from each other in the type of differential equations used (e.g. diffusion systems or reaction diffusion systems). As with network models, most such models use brain connectivity estimates from imaging as a substrate for mediating prion-like protein spreading. However, the differential structure of the models potentially introduces additional temporal complexity: whereas network models produce a timeline corresponding to a fixed pattern of pathology that simply increases steadily in intensity over time, dynamical systems models can produce patterns that change over time, i.e. pathology levels in one region may overtake those in another (as, for instance, the timeline induced by reaction-diffusion systems). Early dynamical systems models, such as in the network diffusion model shown in **Figure 4 (a**), considered pathology spreading from only a single epicentre — enforcing "prion-like" spreading by means of simple linear diffusion models between connected regions[116,119,120]. Iturria-Medina et al.[118] combine terms for appearance and clearance of pathogenic proteins with spreading in their epidemic spreading model (**Figure 4(b)**). They demonstrate qualitatively the model's ability to reconstruct both amyloid[118] and tau[125] deposition patterns in Alzheimer's disease. Later models capture more complex mechanisms such as saturation of protein accumulation[92,115,117,121–124] by incorporating ideas from reaction-diffusion processes. Such models show compelling prediction of pathology timelines in Alzheimer's disease, consistent with observations, e.g., that tau pathology directly precedes atrophy and predicts its topography[126]. Recent approaches also use imaging techniques to inform models in ways that go beyond structural/functional connectivity networks. For example, Weickenmeier et al.[92,121] (**Figure 4(c)**) use a fibre-orientation map from diffusion tensor MRI as a substrate for their reaction-diffusion model of pathology spreading, rather than using derived estimates of brain connectivity.

Most dynamical system models assume a single epicentre common to all patients, although the heterogeneity of neurodegenerative conditions suggests the initial location of pathology appearance may vary. Moreover, appearance may not be limited to a single location or region. Garbarino and Lorenzi[115,122] acknowledge that pathology epicentre likely varies among individuals and personalise individual epicentres to a measured early time-point. Similarly, Torok et al.[127] invert the network diffusion model[116] to wind back the clock from patterns observed in individual patients and show that the most likely single epicentre varies substantially across individuals. Vogel et al.[125] used the epidemic spreading model from[118] to identify distinct epicentres corresponding to each of four subtypes of tau accumulation in Alzheimer's disease, but constrain investigation to a single epicentre region for each

subtype. Identification/optimisation of multi-region epicentres remains a compelling challenge for future work.

While all pathophysiological models describe disease timelines, inference and evaluation of pathophysiological models tends to compare only end-stage predictions of pathology with observed late-stage pathology patterns from individual patients or cohorts. This is reasonable for simple network models where time only increases the intensity of a consistent pattern. However, for more complex dynamical systems models, the temporal evolution of the pattern contains rich information for parameter estimation and model evaluation. Recent studies[95,96,115,117,122,124,128] have moved towards learning pathophysiological models based on their ability to predict the full timeline of the disease rather than just late-stage pathology. This can be achieved by using phenomenological disease progression models as the observed timelines that pathophysiological models aim to capture[115,122].

Mechanistic combinations

Most applications of network models have focused on evaluating the ability of each individual network-metric to explain observed data with the goal of identifying the single mechanism that best explains observed pathology patterns. Similarly, dynamical system models have mostly enforced the prion-like spread hypothesis via network proximity. Both network and dynamical systems models can be extended to model mechanistic combinations, aiming to infer a disease timeline that depends on the relative contribution of multiple mechanisms. Garbarino et al.[94] (see **Figure 4(d)**) suggest that a combination of hypothetical mechanisms likely contributes to timelines observed in any particular disease or individual. Mathematically, the solve a constrained regression problem to infer a set of network metric weightings that combine to define a disease-specific, or even individual, "topological profile" that represents a combination of hypothetical appearance/spreading mechanisms. Iturria-Medina et al.[128–130] explicitly modelled multiple interacting processes (amyloid pathology; atrophy; vascular pathology; glucose metabolism) within a dynamical multimodal network model, mathematically based on control theory of dynamical systems. Fitting the model to patient data suggests vascular dysregulation as an early instigating event in the Alzheimer's pathological cascade. Weickenmeier et al.[121] coupled a reaction-diffusion dynamical system model for the propagation and accumulation of toxic proteins with a mechanical atrophy model for toxin protein-indeced atrophy and showed that the resulting model reproduces typical protein deposition and atrophy patterns found in neurodegenerative diseases. More recently, Lee et al.[101] proposed a network flow-based connectivity model for the mechanisms of interaction of amyloid and tau in Alzheimer's disease, which recapitulates the topographical dissimilarity between early amyloid and tau deposition. He et al.[131] couple various candidate models of appearance and propagation to identify profiles similar to Garbarino et al.[94] and group subjects by similar profile.

Applications

Data-driven disease progression models offer unique potential for impact in fundamental disease understanding and clinical applications. **Figure 5** shows highlights from the literature, but this section discusses both past work and future aspirations.

Biological insight and novel treatment strategies

Early motivators for data-driven disease progression models included understanding the sequence of changes in a neurodegenerative condition, particularly Alzheimer's disease. Such biological insight can inform treatment strategies and have clinical impact. Work to date has generated knowledge regarding how sequences vary among conditions, how

trajectories of multi-modal biomarkers relate on a common timeline, and data-driven verification of qualitative hypothetical models.

Pathology accumulation timescales. Differential equation models (**Figure 5A**) were instrumental in understanding timescales via trajectories of cognitive decline[51,132,133] and pathology accumulation in sporadic[28,52–54,134] and familial[135] Alzheimer's disease. In particular, differential equation models[52,53] (**Figure 5A**) provided the first quantitative estimates of the timescales over which β-amyloid accumulates in Alzheimer's disease, estimating up to 19 years for the transition from early abnormality to a level typical of symptomatic patients.

Quantitative support for hypothetical models. Event-based models of sporadic[38,43] and familial[135] Alzheimer's disease provided quantitative support for hypothetical models[15,136] of the Alzheimer's cascade, along with patient staging tools (**Figure 5C**). They confirmed that accumulation of β-amyloid and tau (quantified using CSF or PET imaging) precede brain atrophy (from MRI) and cognitive decline. Event-based models have further been used to demonstrate predence of new biomarkers to existing ones[137]. Similarly, self-modelling regression approaches[32,61–63,66,70,74,138,139] among others have produced data-driven biomarker trajectories (and staging tools) for Alzheimer's disease[61], Parkinson's disease[50,140], and Huntington's disease[141] that broadly reflect the ordering of changes in hypothetical and anecdotal models, although the precise trajectories of change vary among methods and data sets.

Fine-grained intra-modality models. Fine-grained insights can be yielded from intra-modality models, e.g., regional imaging biomarkers or sets of cognitive test scores. Compelling early image-based results in familial disease[37] led to applications in a wide variety of sporadic and familial diseases, e.g., Huntington's disease[142], Parkinson's dementia[143], progressive multiple sclerosis[144,145], frontotemporal dementia[40,146], amyotrophic lateral sclerosis[147–149], progressive supranuclear palsy[150], and Creutzfeldt-Jakob disease[151]. Event-based models of cognitive tests[39,41] show unique insight into the ordering in which specific cognitive abilities decline, which is particularly useful for understanding atypical, rarer dementias such as posterior cortical atrophy[152].

Subtype characterisation. Subtype models offer new understanding of heterogeneity across the full disease timeline. Applications to date are mostly within-modality, including: highlighting distinct patterns of brain atrophy accumulation in Alzheimer's disease[40,153], frontotemporal dementia[40,146], multiple sclerosis[144], and corticobasal syndrome/progressive supranuclear palsy[154]; patterns of tau and amyloid accumulation in Alzheimer's disease[155,156] and proteomic subtypes of Alzheimer's disease[157]. Image-based data-driven disease subtypes are predictive of genotype in frontotemporal dementia[40], cognitive decline in Alzheimer's disease[155] (**Figure 5B** middle), treatment response in multiple sclerosis[144] (**Figure 5B** upper) and may even find value in pre-symptomatic disease[40,158] (**Figure 5B** lower). Subtype models have been applied to neuropathological ratings data to develop a novel data-driven staging system for TDP-43 pathologies[159].

Influence of disease risk factors. Phenomenological models also offer potential new understanding of how various risk factors influence disease onset and manifestation. For example, Vogel et al.[155] showed that rates of disease progression vary among tau subtypes and Young et al.[160] showed that particular Alzheimer's subtypes associate with cardiovascular/diabetes risk factors. Young et al.[40] also showed that image-based subtypes in genetic frontotemporal dementia broadly align with mutation groups (providing important

validation of the algorithm), but additionally that the *c9orf72* mutation group divides into two distinct manifestations of brain atrophy pattern, as does the *MAPT* mutation group[146], which can be linked to specific mutations in the *MAPT* gene. Those models typically look at risk factor association *post hoc* once the model has been constructed. In contrast, the spatiotemporal model of Koval et al.[70] jointly learns the progression of Alzheimer's disease conditioned on a variety of risk-factor levels to provide a comprehensive picture of how risk factors affect progression in Alzheimer's disease, including earlier and accelerated cognitive decline in women.

Mechanistic insight. Pathophysiological models aim to understand mechanisms of disease. The notion that computational models trained on collections of macroscopic images might reveal which microscopic/molecular processes are at play is compelling. However, experiments to date are simplistic and must be interpreted with caution. The prion hypothesis[102,105,106] emerges strongly from many publications as an essential component for pathophysiological models to explain patient data sets. It underpins the demonstration in Vogel et al.[125] that difference in epicentre is the key difference among manifestations of tau pathology in AD. In-vitro experiments[161] showing that pathological proteins traverse white matter connections locally further support the prion hypothesis, but whether that process scales up to longer ranges and significantly influences whole-brain patterns of pathology remains an open question. Many dynamical systems models[115–120,122,125] assume prion-like spreading and convincingly recapitulate pathology patterns, but others, e.g.,[92,121] use somewhat different spreading models yet also recover convincing patterns. Future frameworks for statistically rigorous model comparison and evaluation of evidence, coupled with in-vitro and in-vivo validation, will be needed to move such techniques from scientific curiosity to serious tools for disease understanding.

Hypothetical treatment strategies. Pathophysiological models can also inform treatment strategies, e.g. by being suggestive of when to intervene and on which biomarker target. Iturria-Medina et al.'s multifactorial differential equation model[128] used simulations to suggest that multi-domain interventions would be most effective. The same group developed a pathophysiological model for estimating clusters of pseudo-temporal cumulative molecular alterations from omics data[58,162] that suggested individualised genetic targets for therapies. Sanz Perl et al.[163] developed a whole-brain perturbational model enabling in silico testing of brain stimulation protocols. While this is exciting, much validation and model refinement work is required to facilitate translation of suggested hypothetical treatment strategies.

Clinical applications and trials

Ultimately, data-driven disease progression models provide quantitative temporal and/or subtype information that can improve individual-level decision-making. This extra information can be used for model-based stratification — or covarying short of stratification — in clinical applications and trials to identify subgroups of responders or enrich cohorts.

Temporal stratification. Model-based *temporal stratification* may help in the design and interpretation of clinical trials. **Figure 5C** shows a multimodal event-based model of sporadic Alzheimer's disease providing fine-grained staging of observational ADNI data[38] (left), and a *post hoc* prognostic enrichment application of event-based modelling[164] (right) on data from the MCI clinical trial[165] where a subgroup of late-stage participants showed superior treatment response. More generally, simulation studies have predicted that fine-grained temporal stratification using data-driven disease progression models can enrich clinical trials,

e.g., in Alzheimer's disease[70,139,166], Huntington's disease[141], and genetic frontotemporal dementia[167].

Subtype stratification. Model-based *subtype stratification* has demonstrated both prognostic and predictive enrichment of clinical trials. **Figure 5B** shows subtyping using SuStaIn[40] being predictive of cognitive decline in Alzheimer's disease[155]. SuStaIn has also been shown to be predictive of treatment response in multiple sclerosis[144] and to have strong subtype assignment for some individuals even in the preclinical (cognitively normal) phase of Alzheimer's disease[40]. Shand et al.[168] used SuStaIn to predict subtype-specific heterogeneity in preclinical Alzheimer's disease cognitive decline in the A4 trial[169], which may be a contributing factor to the subsequent null/negative result[170].

Summary and Future Directions

Summary

Data-driven disease progression models are a novel set of statistical and computational tools for estimating data-driven disease timelines from cross-sectional and short-term longitudinal data, broadly segregating into phenomenological and pathophysiological models. Both phenomenological models and pathophysiological models comprise techniques that estimate disease timelines with an arbitrary timescale based on purely cross-sectional data or with an absolute timescale by incorporating short-term longitudinal data. Across phenomenological models and pathophysiological models, recent developments have focussed on increased personalisation of disease timelines, either through modelling of disease subtypes, individualised trajectories, or mechanistic combinations. Currently a range of phenomenological and pathophysiological models exist that have similar outputs but are subtly different in their mathematical structure; direct comparison of models is necessary to characterise differences in model performance.

Phenomenological and pathophysiological models have evolved separately with distinct aims; phenomenological models have direct application for staging, stratification, prognostication, differential diagnosis and other classification and prediction tasks, whilst pathophysiological models evaluate the evidence for different candidate disease mechanisms or infer key parameters that govern disease mechanisms. Phenomenological models are more technologically mature than pathophysiological models; many contributions in the literature include open-source software tools (for a summary see[171]), and have been replicated across multiple datasets. Comparatively, pathophysiological models are in earlier stages of technological development and as such their outputs should be treated with greater caution. The ability to perform classification and prediction tasks and greater technological maturity of phenomenological models gives them direct applicability to clinical trial and healthcare settings. However, pathophysiological models offer more detailed biological insights that can inform treatment and prevention strategies and may ultimately prove to be a more powerful framework for integrating data across scales to build individualised models, thus achieving greater predictive power as well as providing tailored biological insights.

Future technical directions

Current data-driven disease progression models have a range of limitations, highlighting key areas for future technical development.

Feature learning. Most current models depend on a pre-specified set of input features, limiting the richness with which disease timelines can be inferred, and potentially missing relationships between, e.g., distal brain areas. Feature learning, i.e., simultaneous estimation of features and their timelines could replace pre-defined disease features (e.g. regional brain volumes, pre-selected genetic variants) with more salient, potentially diffuse/multi-modal, feature sets to define trajectories. Marinescu et al.[75] provides an early demonstration of how this might be achieved and deep learning approaches such as Yang et al.[172] hold significant promise.

Incorporating treatment effects. As treatments become available for neurodegenerative diseases[173,174], the ability to model and predict treatment effects becomes vital. Future work could relax the assumption of monotonic trajectories to model treatment response, either through careful consideration of the use of longitudinal data or interventional studies to exploit and learn causal effects e.g.[175].

Omics-based models. Integration of omics information[57–59] could increase the biological insight provided by both phenomenological and pathophysiological models. To date, associations between disease progression models and genetics have predominantly been identified post-hoc, e.g.[146,176]. Future models could directly consider genetic variants as fixed effects to infer how genetic risk factors influence progression patterns or speed of progression. More complex models may also be possible, for example using single cell omics data to compare hypotheses about disease mechanisms using data from a range of different cell types.

Integrative models across scales. Models that integrate multi-omic data across scales hold significant promise. Integration with pseudotime methods developed by the single-cell transcriptomics community could offer a platform for development of such techniques, as demonstrated by Iturria-Medina et al.[57–59]. Models that describe timelines associating functional and anatomic changes with deficits in cognitive subdomains may also be possible by combining brain-behaviour models[177] with disease progression models.

Age effects. Age is the biggest risk factor for dementia and the aging process may interact with neurodegenerative pathologies. The integration of disease progression modelling with advanced models of aging such as ‘brain age’ models[178–180] could offer two key benefits: (i) gaining mechanistic insights into the interaction between the aging process and neurodegenerative disease and (ii) enabling better discrimination of health and disease through normative modelling[181].

Multi-morbidity. Multi-morbidity — the co-existence of multiple chronic health conditions — is common among elderly populations. Co-morbid conditions could be modelled as a fixed effects (or interaction terms) in disease progression models.

Mixed pathology. Mixed pathology — the presence of pathologies consistent with multiple neurodegenerative diseases — is highly prevalent in elderly populations. Current disease progression models fail to isolate the effects and interactions of each pathology. Future models could consider each pathology as a separate component[182], enabling the effects of individual pathologies on non-specific downstream biomarkers such as brain volume loss and cognition to be disentangled.

Wider mechanisms. Pathophysiological models could be adapted to consider a wider range of disease mechanisms. Examples include: metabolic alterations, e.g. deficiencies in glucose uptake[183], systemic mitochondrial dysfunction, oxidative damage and lipid

metabolism[184,185]; dysfunction of cerebral vasculature[186–188]; neuroinflammation / glial cell activation[189–191].

Broader neurological and non-neurological applications. Disease progression models are increasingly being applied to a wider range of neurological and non-neurological conditions beyond primary neurodegenerative diseases. Examples include Multiple Sclerosis[144,192,193], Schizophrenia[194], depression[195], osteoarthritis[196], eye[197], and lung disease[198]. Tailoring disease progression models to these individual application areas represents a key area for future model development.

Spectra vs subtypes. Current subtype models assume distinct groups and will "discover" subtypes even when variation is more spectral in nature. Recent work in manifold learning aims to discriminate these scenarios[199] using continuous representations and could be used in combination with disease progression modelling to better understand whether subtypes are distinct entities or represent points in a landscape of variation.

Evaluation of models. Thorough evaluation of models is key to realising the translational potential of disease progression models and offers important future challenges. The vast majority of Alzheimer's disease progression models are built using the publicly available Alzheimer's Disease Neuroimaging Initiative dataset, but external validation is increasing[155,200–202] with greater availability of external datasets e.g., [203]. While evaluation is challenging in the absence of a ground truth, strategies include generating a synthetic ground truth (e.g., simulated disease timelines[70,204]), comparing model outputs with known biological characteristics (such as genetics[40]), assessing predictive performance and longitudinal consistency (in follow-up data[135]), and external validation on the ever-increasing volume of data available[70,92,116,122,144,200–202,205]. One key future direction is to benchmark models against one another, to better understand which components of disease progression modelling are the most beneficial in different applications[70,204]. This will be advanced through further community challenges, such as The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge (https://tadpole.grand-challenge.org)[206,207], CADDementia[208], and others[209]. Pragmatic investigations should include minimum data requirements for model fitting (and downstream insight) to inform the design of resource-efficient and patient-friendly study protocols in both observational and interventional studies. Additionally, model performance in differential diagnosis applications remains under-explored, in part due to protocol differences between disease studies.

Conclusion and Perspectives

The rapidly evolving area of data-driven disease progression modelling offers potentially transformational opportunities for applications along two key paths: i) biological understanding of disease to inform strategies for prevention or intervention; and ii) applications in clinical trials and healthcare. The range of applications in clinical trials and healthcare are diverse, including clinical trial design, use in healthcare settings for patient diagnosis and prognosis, and use in population health planning and policy making. We will continue to learn about the diversity, interaction, pathways, and causes of neurodegenerative diseases through coupled advances in understanding of the fundamental biology of these diseases, and model precision. The advent of disease modifying therapies provides an additional means of refining the models and assessing the impact of manipulating the pathological process.

Data-driven disease progression models provide a platform to use human knowledge to guide machine-based discovery. This will require a transition from current clinical practice where the influence of disease progression models is currently secondary, i.e., model

insights guiding and informing systems that are designed by humans. For example, the current proposed research A/T/N staging system[210] for Alzheimer's disease has the anatomy of an event-based model with subtyping (multiple series of stages each defined by the appearance of a symptom or abnormal marker) but was constructed based on heuristic human insight rather than being directly data-driven. Similarly, clinical trials have not yet used disease progression models directly, with data-driven insights only indirectly informing on trial design. In future, disease progression models could directly be used for stratification, covarying short of stratification, or prediction of outcomes to inform on optimal clinical trial designs.

Realising the potential of data-driven disease progression modelling across these broad application areas involves diverse efforts: further technical development of the models themselves; continued data collection and validation activities; engagement with the wide variety of stakeholders (clinicians, drug developers, medical policy makers, patients, etc.) to ensure understanding and enable uptake. Linking together specialists in different disciplines will be crucial to enabling the broad range of potential applications. High-quality data for training and validation of models is key and may ultimately require coordinated efforts between methods developers and clinicians or pharmaceutical companies to collect tailored datasets. Long term we envision data-driven disease progression modelling as part of a new computational medicine paradigm that brings the kinds of techniques we discuss here centre stage in the diagnosis, staging and prognostication of patients with neurodegenerative diseases, and in informing therapeutic decision making.

Competing Interests

NPO consults for TheraPanacea (FR) and Queen Square Analytics (UK). FB and DCA hold an equity stake in Queen Square Analytics (UK). FB is a steering committee or Data Safety Monitoring Board member for Biogen, Merck, ATRI/ACTC and Prothena. He is a consultant for Roche, Celltrion, Rewind Therapeutics, Merck, IXICO, Jansen, Combinostics and has research agreements with Merck, Biogen, GE Healthcare, Roche. All other authors have no competing interests to declare.

Acknowledgements

ALY was supported by a Skills Development Fellowship (MR/T027800/1) from the Medical Research Council and a Career Development Award from the Wellcome Trust [227341/Z/23/Z]. NPO acknowledges funding from his UKRI Future Leaders Fellowship (MR/S03546X/1), and the Early Detection of Alzheimer's Disease Subtypes project (E-DADS; EU JPND, MR/T046422/1). SG acknowledges financial contribution from the National Group of Scientific Computing with the INdAM–GNCS Project (CUP_E53C22001930001) - Computational methods for modelling neurodegenerative diseases progression. SG acknowledges the support of NEXTGENERATIONEU (NGEU) and fundings by the Italian Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) – A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022). SG acknowledges financial contribution from the Italian Ministry of Health, with the project NeuroArtP3 (NET-2018-12366666) - Artificial intelligence of imaging and clinical neurological data for predictive, preventive, and personalized medicine. FB is supported by the NIHR Biomedical Research Centre at UCLH. JMS acknowledges the support of the National Institute for Health Research University College London Hospitals Biomedical Research Centre, Wolfson Foundation, Alzheimer’s Research UK, Brain Research UK, Weston Brain Institute, Medical Research Council, British Heart Foundation, and Alzheimer’s Association.
The Wellcome Trust [221915/Z/20/Z], JPND / MRC grant MR/T046422/1, and the NIHR UCLH Biomedical Research Centre support DA’s work on this topic. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 666992.

References

1. Bateman, R. J. *et al.* Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* **367**, 795–804 (2012).
2. Savica, R., Rocca, W. A. & Ahlskog, J. E. When does Parkinson disease start? *Arch. Neurol.* **67**, 798–801 (2010).
3. Rohrer, J. D. *et al.* Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: A cross-sectional analysis. *Lancet Neurol.* **14**, 253–262 (2015).
4. Paulsen, J. S. *et al.* Detection of Huntington's disease decades before diagnosis: The Predict-HD study. *J. Neurol. Neurosurg. Psychiatry* **79**, 874–880 (2008).
5. Hansson, O. Biomarkers for neurodegenerative diseases. *Nat. Med.* **27**, 954–963 (2021).
6. Makhani, N. & Tremlett, H. The multiple sclerosis prodrome. *Nat. Rev. Neurol.* **17**, 515–521 (2021).
7. Lange, P. *et al.* Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N. Engl. J. Med.* **373**, 111–122 (2015).
8. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
9. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791–800 (2002).
10. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479 (1991).
11. Nelson, P. T. *et al.* Limbic-predominant age-related TDP-43 encephalopathy (LATE): Consensus working group report. *Brain* **142**, 1503–1527 (2019).
12. Josephs, K. A. *et al.* Staging TDP-43 pathology in Alzheimer's disease. *Acta Neuropathol.* **127**, 441–450 (2014).
13. Brettschneider, J., Del Tredici, K., Lee, V. M. Y. & Trojanowski, J. Q. Spreading of pathology in neurodegenerative diseases: A focus on human studies. *Nat. Rev. Neurosci.* **16**, 109–120 (2015).
14. Braak, H. *et al.* Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
15. Jack, C. R. *et al.* Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol.* **9**, 119–28 (2010).
16. Jack, C. R. *et al.* Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* **12**, 207–16 (2013).
17. Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P. & Thompson, P. M. The clinical use of structural MRI in Alzheimer disease. *Nat. Rev. Neurol.* **6**, 67–77 (2010).
18. Frisoni, G. B. *et al.* The probabilistic model of Alzheimer disease: the amyloid hypothesis revised. *Nat. Rev. Neurosci.* **23**, 53–66 (2022).
19. Fleisher, A. S. *et al.* Associations between biomarkers and age in the presenilin 1 E280A autosomal dominant Alzheimer disease kindred: a cross-sectional study. *JAMA Neurol.* **72**, 316–24 (2015).
20. Scahill, R. I., Schott, J. M., Stevens, J. M., Rossor, M. N. & Fox, N. C. Mapping the evolution of regional atrophy in Alzheimer's disease : Unbiased analysis of fluid-registered serial MRI. *Proc. Natl. Acad. Sci.* **99**, 1–5 (2002).
21. Ridha, B. H. *et al.* Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *Lancet Neurol.* **5**, 828–834 (2006).
22. Fox, N. C. & Schott, J. M. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *Lancet* **363**, 392–394 (2004).
23. Jack, C. R. *et al.* Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease. *Brain* **132**, 1355–65 (2009).

24. Lo, R. Y. *et al.* Longitudinal change of biomarkers in cognitive decline. *Arch. Neurol.* **68**, 1257–66 (2011).
25. Landau, S. M. *et al.* Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Ann. Neurol.* **72**, 578–86 (2012).
26. Jack, C. R. *et al.* Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. *Arch. Neurol.* **69**, 856–67 (2012).
27. Caroli, A. & Frisoni, G. B. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. *Neurobiol. Aging* **31**, 1263–74 (2010).
28. Sabuncu, M. R. *et al.* The dynamics of cortical and hippocampal atrophy in Alzheimer disease. *Arch. Neurol.* **68**, 1040–8 (2011).
29. Jack, C. R. *et al.* Evidence for ordering of Alzheimer disease biomarkers. *Arch. Neurol.* **68**, 1526–35 (2011).
30. Ryman, D. C. *et al.* Symptom onset in autosomal dominant Alzheimer disease A systematic review and meta-analysis. *Neurology* **83**, 253–260 (2014).
31. Schmidt-Richberg, A. *et al.* Learning Biomarker Models for Progression Estimation of Alzheimer's Disease. *PLoS One* **11**, e0153040 (2016).
32. Guerrero, R. *et al.* Instantiated mixed effects modeling of Alzheimer's disease markers. *Neuroimage* (2016). doi:10.1016/j.neuroimage.2016.06.049
33. Poulakis, K. *et al.* Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. *Nat. Commun.* **13**, (2022).
34. Buchhave, P. *et al.* Cerebrospinal fluid levels of β-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. *Arch. Gen. Psychiatry* **69**, 98–106 (2012).
35. Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease A systematic review and meta-analysis. *Neurology* **94**, (2020).
36. Habes, M. *et al.* Disentangling Heterogeneity in Alzheimer's Disease and Related Dementias Using Data- Driven Methods. *Biol. Psychiatry* **88**, 70–82 (2020).
37. Fonteijn, H. M. *et al.* An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *Neuroimage* **60**, 1880–9 (2012).
38. Young, A. L. *et al.* A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* **137**, 2564–2577 (2014).
39. Firth, N. C. *et al.* Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression. *Alzheimer's Dement.* **16**, 965–973 (2020).
40. Young, A. L. *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat. Commun.* **9**, 4273 (2018).
41. Young, A. L. *et al.* Ordinal SuStaIn: Subtype and Stage Inference for Clinical Scores, Visual Ratings, and Other Ordinal Data. *Front. Artif. Intell.* **4**, 1–13 (2021).
42. Huang, J. & Alexander, D. Probabilistic Event Cascades for Alzheimer's disease. in *Advances in Neural Information Processing Systems* 3104–3112 (2012).
43. Venkatraghavan, V., Bron, E. E., Niessen, W. J. & Klein, S. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. *Neuroimage* **186**, 518–532 (2019).
44. Tandon, R., Kirkpatrick, A. & Mitchell, C. S. *sEBM: Scaling Event Based Models to Predict Disease Progression via Implicit Biomarker Selection and Clustering*. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **13939 LNCS**, (Springer Nature Switzerland, 2023).
45. Parker, C., Oxtoby, N., Alexander, D., Zhang, H. & Initiative, the A. D. N. S-EBM: Generalising event-based modelling of disease progression for simultaneous events. *bioRxiv* 2022.07.10.499471 (2022).
46. Du, J. & Zhou, Y. *Filtered Trajectory Recovery: A Continuous Extension to Event-*

Based Model for Alzheimer's Disease Progression Modeling. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **13939 LNCS**, (Springer Nature Switzerland, 2023).
47. Wijeratne, P. A. & Alexander, D. C. Learning Transition Times in Event Sequences: The Temporal Event-Based Model of Disease Progression. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **12729 LNCS**, 583–595 (2021).
48. Wijeratne, P. A. *et al.* The temporal event-based model: Learning event timelines in progressive diseases. *Imaging Neurosci.* **1**, 1–19 (2023).
49. Severson, K. A. *et al. Personalized Input-Output Hidden Markov Models for Disease Progression Modeling*. *Proceedings of Machine Learning Research* **126**, (2020).
50. Severson, K. A. *et al.* Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning. *Lancet Digit. Heal.* **3**, e555–e564 (2021).
51. Samtani, M. N. *et al.* 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**, 146–61 (2013).
52. Villemagne, V. L. *et al.* Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* **12**, 357–67 (2013).
53. Jack, C. R. *et al.* Brain β-amyloid load approaches a plateau. *Neurology* **80**, 890–6 (2013).
54. Oxtoby, N. P. *et al.* Learning imaging biomarker trajectories from noisy Alzheimer's disease data using a Bayesian multilevel model. in *Bayesian and Graphical Models for Biomedical Imaging* (eds. Simpson, I., Arbel, T., Ribbens, A., Cardoso, M. J. & Precup, D.) **8677**, 85–94 (Springer International Publishing, 2014).
55. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
56. Saint-Jalmes, M. *et al.* Disease progression modelling of Alzheimer's disease using probabilistic principal components analysis. *Neuroimage* **278**, (2023).
57. Iturria-Medina, Y., Khan, A. F., Adewale, Q. & Shirazi, A. H. Blood and brain gene expression trajectories mirror neuropathology and clinical deterioration in neurodegeneration. *Brain* **143**, 661–673 (2020).
58. Iturria-Medina, Y. *et al.* Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of Alzheimer's disease progression and heterogeneity. *Sci. Adv.* **8**, 1–19 (2022).
59. McCarthy, J. *et al.* Data-driven staging of genetic frontotemporal dementia using multi-modal MRI. *Hum. Brain Mapp.* **43**, 1821–1835 (2022).
60. Jedynak, B. M. *et al.* A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort. *Neuroimage* **63**, 1478–86 (2012).
61. Donohue, M. C. *et al.* Estimating long-term multivariate progression from short-term data. *Alzheimer's Dement.* **10**, S400–S410 (2014).
62. Li, D., Iddi, S., Thompson, W. K. & Donohue, M. C. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Stat. Methods Med. Res.* **28**, 835–845 (2019).
63. Lorenzi, M., Filippone, M., Frisoni, G. B., Alexander, D. C. & Ourselin, S. Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease. *NeuroImage* **190**, 56–68 (2019).
64. Raket, L. L. Statistical Disease Progression Modeling in Alzheimer Disease. *Front. Big Data* **3**, 1–18 (2020).
65. Schiratti, J. B., Allassonnière, S., Colliot, O. & Durrleman, S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. in *Advances in Neural Information Processing Systems* **2015-Janua**, 2404–2412 (2015).
66. Schiratti, J. B., Allassonnière, S., Routier, A., Colliot, O. & Durrleman, S. A mixed-

effects model with time reparametrization for longitudinal univariate manifold-valued data. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9123**, 564–575 (2015).
67. Schiratti, J., Allassonniere, S., Colliot, O. & Durrleman, S. Mixed-effects model for the spatiotemporal analysis of longitudinal manifold-valued data. *Proc. 5th MICCAI Work. Math. Found. Comput. tional Anat.* 1–12 (2015).
68. Schiratti, J. B., Allassonnière, S., Colliot, O. & Durrleman, S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *J. Mach. Learn. Res.* **18**, 1–33 (2017).
69. Louis, M., Couronné, R., Koval, I., Charlier, B. & Durrleman, S. Riemannian Geometry Learning for Disease Progression Modelling. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **11492 LNCS**, 542–553 (2019).
70. Koval, I. *et al.* AD Course Map charts Alzheimer's disease progression. *Sci. Rep.* **11**, 1–16 (2021).
71. Gruffaz, S., Poulet, P. E., Maheux, E., Jedynak, B. & Durrleman, S. Learning Riemannian metric for disease progression modeling. *Adv. Neural Inf. Process. Syst.* **28**, 23780–23792 (2021).
72. Lorenzi, M., Pennec, X., Frisoni, G. B. & Ayache, N. Disentangling normal aging from Alzheimer's disease in structural magnetic resonance images. *Neurobiol. Aging* **36**, S42–S52 (2015).
73. Durrleman, S. *et al.* Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *Int. J. Comput. Vis.* **103**, 22–59 (2013).
74. Bilgel, M., Prince, J. L., Wong, D. F., Resnick, S. M. & Jedynak, B. M. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage* **134**, 658–670 (2016).
75. Marinescu, R. V. *et al.* DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders. *Neuroimage* **192**, 166–177 (2019).
76. Abi Nader, C., Ayache, N., Robert, P. & Lorenzi, M. Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data. *Neuroimage* **205**, (2020).
77. Durrleman, S., Pennec, X., Trouve, A., Gerig, G. & Ayache, N. Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. *Int. Conf. Med. Image Comput. Comput. Interv.* 297–304 (2009).
78. Aksman, L. M. *et al.* pySuStaIn: A Python implementation of the Subtype and Stage Inference algorithm. *SoftwareX* **16**, 100811 (2021).
79. Young, A. L., Aksman, L. M., Alexander, D. C. & Wijeratne, P. A. *Subtype and Stage Inference with Timescales*. **2**, (Springer Nature Switzerland, 2023).
80. Chen, I. Y., Krishnan, R. G. & Sontag, D. Clustering Interval-Censored Time-Series for Disease Phenotyping. *AAAI Conf.* **1**, (2022).
81. Poulet, P. E. & Durrleman, S. Mixture Modeling for Identifying Subtypes in Disease Course Mapping. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **12729 LNCS**, 571–582 (Springer International Publishing, 2021).
82. Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L. & Seeley, W. W. Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron* **73**, 1216–27 (2012).
83. Altmann, A. *et al.* Analysis of brain atrophy and local gene expression in genetic frontotemporal dementia. *Brain Commun.* **2**, 1–13 (2020).
84. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron* **62**, 42–52 (2009).
85. Saxena, S. & Caroni, P. Selective Neuronal Vulnerability in Neurodegenerative Diseases: From Stressor Thresholds to Degeneration. *Neuron* **71**, 35–48 (2011).
86. de Haan, W., Mott, K., van Straaten, E. C. W., Scheltens, P. & Stam, C. J. Activity

Dependent Degeneration Explains Hub Vulnerability in Alzheimer's Disease. *PLoS Comput. Biol.* **8**, (2012).
87. Fu, H., Hardy, J. & Duff, K. E. Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1350–1358 (2018).
88. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).
89. Bateman, R. J. *et al.* Human amyloid-β synthesis and clearance rates as measured in cerebrospinal fluid in vivo. *Nat. Med.* **12**, 856–861 (2006).
90. Clavaguera, F. *et al.* 'Prion-Like' templated misfolding in tauopathies. in *Brain Pathology* **23**, 342–349 (2013).
91. Bourdenx, M. *et al.* Progress in Neurobiology Protein aggregation and neurodegeneration in prototypical neurodegenerative diseases : Examples of amyloidopathies , tauopathies and synucleinopathies. *Prog. Neurobiol.* **155**, 171–193 (2017).
92. Weickenmeier, J., Jucker, M., Goriely, A. & Kuhl, E. Journal of the Mechanics and Physics of Solids A physics-based model explains the prion-like features of neurodegeneration in Alzheimer ' s disease , Parkinson ' s disease , and amyotrophic lateral sclerosis. *J. Mech. Phys. Solids* **124**, 264–281 (2019).
93. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature Neuroscience* **21**, 1332–1340 (2018).
94. Garbarino, S. *et al.* Differences in topological progression profile among neurodegenerative diseases from imaging data. *Elife* **8**, 1–27 (2019).
95. Oxtoby, N. P. *et al.* Data-driven sequence of changes to anatomical brain connectivity in sporadic Alzheimer's disease. *Front. Neurol.* **8**, 1–11 (2017).
96. Cope, T. E. *et al.* Tau burden and the functional connectome in Alzheimer's disease and progressive supranuclear palsy. *Brain* **141**, 550–567 (2018).
97. Brown, J. A. *et al.* Patient-Tailored, Connectivity-Based Forecasts of Spreading Brain Atrophy. *Neuron* **104**, 856-868.e5 (2019).
98. Sintini, I. *et al.* Tau and Amyloid Relationships with Resting-state Functional Connectivity in Atypical Alzheimer's Disease. *Cereb. Cortex* **31**, 1693–1706 (2021).
99. Franzmeier, N. *et al.* Functional brain architecture is associated with the rate of tau accumulation in Alzheimer's disease. *Nat. Commun.* **11**, 1–17 (2020).
100. Franzmeier, N. *et al.* Tau deposition patterns are associated with functional connectivity in primary tauopathies. *Nat. Commun.* **13**, 1–18 (2022).
101. Lee, W. J. *et al.* Regional Aβ-tau interactions promote onset and acceleration of Alzheimer's disease tau spreading. *Neuron* **110**, 1932-1943.e5 (2022).
102. Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45–51 (2013).
103. Tijms, B. M. *et al.* Single-Subject Grey Matter Graphs in Alzheimer's Disease. *PLoS One* **8**, 1–9 (2013).
104. Pelkmans, W. *et al.* Grey matter network markers identify individuals with prodromal Alzheimer's disease who will show rapid clinical decline. *Brain Commun.* **4**, 1–9 (2022).
105. Frost, B. & Diamond, M. I. Prion-like mechanisms in neurodegenerative diseases. *Nat. Rev. Neurosci.* **11**, 155–159 (2010).
106. Prusiner, S. B. Some Speculations about Prions, Amyloid and Alzheimer's disease. *Nejm* **310**, 661–663 (1984).
107. Zott, B. *et al.* A vicious cycle of β amyloid−dependent neuronal hyperactivation. *Science (80-.).* **365**, 559–565 (2019).
108. Appel, S. H. A unifying hypothesis for the cause of amyotrophic lateral sclerosis, parkinsonism, and alzheimer disease. *Ann. Neurol.* **10**, 499–505 (1981).
109. Salehi, A. *et al.* Increased App Expression in a Mouse Model of Down ' s Syndrome Disrupts NGF Transport and Causes Cholinergic Neuron Degeneration. *Neuron* 29–42 (2006). doi:10.1016/j.neuron.2006.05.022
110. Cioli, C., Abdi, H., Beaton, D., Burnod, Y. & Mesmoudi, S. Differences in human

cortical gene expression match the temporal properties of large-scale functional networks. *PLoS One* **9**, 1–28 (2014).
111. Rittman, T. *et al.* Regional expression of the MAPT gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy. *Neurobiol. Aging* **48**, 153–160 (2016).
112. Leng, K. *et al.* Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **24**, 276–287 (2021).
113. Damoiseaux, J. S. & Greicius, M. D. Greater than the sum of its parts: A review of studies combining structural connectivity and resting-state functional connectivity. *Brain Struct. Funct.* **213**, 525–533 (2009).
114. Chennu, S. *et al.* Brain networks predict metabolism, diagnosis and prognosis at the bedside in disorders of consciousness. *Brain* **140**, 2120–2132 (2017).
115. Garbarino, S. & Lorenzi, M. Investigating hypotheses of neurodegeneration by learning dynamical systems of protein propagation in the brain. *Neuroimage* **235**, 117980 (2021).
116. Raj, A., Kuceyeski, A. & Weiner, M. A network diffusion model of disease progression in dementia. *Neuron* **73**, 1204–15 (2012).
117. Schäfer, A., Peirlinck, M., Linka, K. & Kuhl, E. Bayesian Physics-Based Modeling of Tau Propagation in Alzheimer's Disease. *Front. Physiol.* **12**, 1–12 (2021).
118. Iturria-Medina, Y., Sotero, R. C., Toussaint, P. J. & Evans, A. C. Epidemic Spreading Model to Characterize Misfolded Proteins Propagation in Aging and Associated Neurodegenerative Disorders. *PLoS Comput. Biol.* **10**, e1003956 (2014).
119. Raj, A. *et al.* Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer's Disease. *Cell Rep.* **10**, 359–369 (2015).
120. Mišić, B. *et al.* Cooperative and Competitive Spreading Dynamics on the Human Connectome. *Neuron* **86**, 1518–1529 (2015).
121. Weickenmeier, J., Kuhl, E. & Goriely, A. Multiphysics of Prionlike Diseases : Progression and Atrophy. *Phys. Rev. Lett.* **121**, 158101 (2018).
122. Garbarino, S. & Lorenzi, M. Modeling and Inference of Spatio-Temporal Protein Dynamics Across Brain Networks. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11492 LNCS**, 57–69 (Springer International Publishing, 2019).
123. Bertsch, M., Franchi, B., Meacci, L., Primicerio, M. & Tesi, M. C. The amyloid cascade hypothesis and Alzheimer's disease: A mathematical model. *Eur. J. Appl. Math.* **32**, 749–768 (2021).
124. Powell, F., Tosun, D. & Raj, A. Network-constrained technique to characterize pathology progression rate in Alzheimer's disease. *Brain Commun.* **3**, 1–16 (2021).
125. Vogel, J. W. *et al.* Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease. *Nat. Commun.* **11**, 2612 (2020).
126. Iaccarino, L. *et al.* NeuroImage : Clinical Local and distant relationships between amyloid , tau and neurodegeneration in Alzheimer ' s Disease. *NeuroImage Clin.* **17**, 452–464 (2018).
127. Torok, J., Maia, P. D., Powell, F., Pandya, S. & Raj, A. A method for inferring regional origins of neurodegeneration. *Brain* **141**, 863–876 (2018).
128. Iturria-Medina, Y., Carbonell, F. M., Sotero, R. C., Chouinard-Decorte, F. & Evans, A. C. Multifactorial causal model of brain (dis)organization and therapeutic intervention: Application to Alzheimer's disease. *Neuroimage* **152**, 60–77 (2017).
129. Iturria-Medina, Y., Sotero, R. C., Toussaint, P. J., Mateos-Perez, J. M. & Evans, A. C. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nat. Commun.* **7**, 11934 (2016).
130. Iturria-Medina, Y., Carbonell, F. M. & Evans, A. C. Multimodal imaging-based therapeutic fingerprints for optimizing personalized interventions: Application to neurodegeneration. *Neuroimage* **179**, 40–50 (2018).
131. He, T. *et al. A coupled-mechanisms modelling framework for neurodegeneration*. **1**, (Springer Nature Switzerland, 2023).

132. Ashford, J. W. & Schmitt, F. A. Modeling the time-course of Alzheimer dementia. *Curr. Psychiatry Rep.* **3**, 20–8 (2001).
133. Gomeni, R. *et al.* Modeling Alzheimer's disease progression using the disease system analysis approach. *Alzheimer's Dement.* **8**, 39–50 (2012).
134. Budgeon, C. A. *et al.* Constructing longitudinal disease progression curves using sparse, short-term individual data with an application to Alzheimer's disease. *Stat. Med.* **36**, 2720–2734 (2017).
135. Oxtoby, N. P. *et al.* Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain* 1–16 (2018). doi:10.1093/brain/awy050
136. Aisen, P. S. *et al.* Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimer's Dement.* **6**, 239–46 (2010).
137. Byrne, L. M. *et al.* Evaluation of mutant huntingtin and neurofilament proteins as potential markers in Huntington's disease. *Sci. Transl. Med.* **10**, (2018).
138. Bilgel, M. & Jedynak, B. M. Predicting time to dementia using a quantitative template of disease progression. *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **11**, 205–215 (2019).
139. Maheux, E. *et al.* Forecasting individual progression trajectories in Alzheimer's disease. *Nat. Commun.* **14**, (2023).
140. Iddi, S. *et al.* Estimating the evolution of disease in the Parkinson's Progression Markers Initiative. *Neurodegener Dis* **18**, 173–190 (2018).
141. Koval, I. *et al.* Forecasting individual progression trajectories in Huntington disease enables more powered clinical trials. *Sci. Rep.* **12**, 1–14 (2022).
142. Wijeratne, P. A. *et al.* An image-based model of brain volume biomarker changes in Huntington's disease. *Ann. Clin. Transl. Neurol.* **5**, 570–582 (2018).
143. Oxtoby, N. P. *et al.* Sequence of clinical and neurodegeneration events in Parkinson's disease progression. *Brain* **144**, 975–988 (2021).
144. Eshaghi, A. *et al.* Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat. Commun.* **12**, 2078 (2021).
145. Dekker, I. *et al.* The sequence of structural, functional and cognitive changes in multiple sclerosis. *NeuroImage Clin.* **29**, 102550 (2021).
146. Young, A. L. *et al.* Characterizing the Clinical Features and Atrophy Patterns of MAPT-Related Frontotemporal Dementia With Disease Progression Modeling. *Neurology* **97**, e941–e952 (2021).
147. Gabel, M. C. *et al.* Evolution of white matter damage in amyotrophic lateral sclerosis. *Ann. Clin. Transl. Neurol.* **7**, 722–732 (2020).
148. Broad, R. J. *et al.* Neurite orientation and dispersion density imaging (NODDI) detects cortical and corticospinal tract degeneration in ALS. *J. Neurol. Neurosurg. Psychiatry* **90**, 404–411 (2019).
149. Wen, J. *et al.* Neurite density is reduced in the presymptomatic phase of C9orf72 disease. *J. Neurol. Neurosurg. Psychiatry* **90**, 387–394 (2019).
150. Scotton, W. J. *et al.* A data-driven model of brain volume changes in progressive supranuclear palsy. *Brain Commun.* **4**, (2022).
151. Pascuzzo, R. *et al.* Prion propagation estimated from brain diffusion MRI is subtype dependent in sporadic Creutzfeldt–Jakob disease. *Acta Neuropathol.* **140**, 169–181 (2020).
152. Firth, N. C. *et al.* Longitudinal neuroanatomical and cognitive progression of posterior cortical atrophy. *Brain* **142**, 2082–2095 (2019).
153. ten Kate, M. *et al.* Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. *Brain* **141**, 3443–3456 (2018).
154. Saito, Y. *et al.* Temporal Progression Patterns of Brain Atrophy in Corticobasal Syndrome and Progressive Supranuclear Palsy Revealed by Subtype and Stage Inference (SuStaIn). *Front. Neurol.* **13**, (2022).
155. Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat. Med.* **27**, 871–881 (2021).
156. Collij, L. E. *et al. Spatial-Temporal Patterns of Amyloid-β Accumulation: A Subtype*

and Stage Inference Model Analysis. *Neurology* **0**, (2022).
157. Tijms, B. M. *et al.* Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain* **143**, 3776–3792 (2020).
158. Chen, H. *et al.* Transferability of Alzheimer's disease progression subtypes to an independent population cohort. *Neuroimage* **271**, 120005 (2023).
159. Young, A. L. *et al.* Data-driven neuropathological staging and subtyping of TDP-43 proteinopathies. *Brain* **146**, 2975–2988 (2023).
160. Young, A. L. *et al.* Genomewide association study of data-driven Alzheimer's disease subtypes. in *Alzheimer's Association International Conferenceo* (2018).
161. Ahmed, Z. *et al.* A novel in vivo model of tau propagation with rapid and progressive neurofibrillary tangle pathology: The pattern of spread is determined by connectivity, not proximity. *Acta Neuropathol.* **127**, 667–683 (2014).
162. Adewale, Q., Khan, A. F., Carbonell, F. & Iturria-Medina, Y. Integrated transcriptomic and neuroimaging brain model decodes biological mechanisms in aging and Alzheimer's disease. *Elife* **10**, 1–22 (2021).
163. Perl, Y. S. *et al.* Model-based whole-brain perturbational landscape of neurodegenerative diseases. *Elife* **12**, 1–25 (2023).
164. Oxtoby, N. P., Shand, C., Cash, D. M., Alexander, D. C. & Barkhof, F. Targeted Screening for Alzheimer's Disease Clinical Trials Using Data-Driven Disease Progression Models. *Front. Artif. Intell.* **5**, 1–9 (2022).
165. Petersen, R. C. *et al.* Vitamin E and Donepezil for the Treatment of Mild Cognitive Impairment. *N. Engl. J. Med.* **352**, 2379–88 (2005).
166. Abi Nader, C., Ayache, N., Frisoni, G. B., Robert, P. & Lorenzi, M. Simulating the outcome of amyloid treatments in Alzheimer's disease from imaging and clinical data. *Brain Commun.* **3**, 1–17 (2021).
167. Staffaroni, A. M. *et al.* Temporal order of clinical and biomarker changes in familial frontotemporal dementia. *Nat. Med.* **28**, 2194–2206 (2022).
168. Shand, C. *et al.* Heterogeneity in Preclinical Alzheimer's Disease Trial Cohort Identified by Image-based Data-Driven Disease Progression Modelling. *medRxiv* 1–19 (2023).
169. Sperling, R. A. *et al.* Association of Factors with Elevated Amyloid Burden in Clinically Normal Older Individuals. *JAMA Neurol.* **77**, 735–745 (2020).
170. Sperling, R. A. *et al.* Trial of Solanezumab in Preclinical Alzheimer's Disease. *N. Engl. J. Med.* (2023). doi:10.1056/nejmoa2305032
171. Oxtoby, N. P. Data-Driven Disease Progression Modelling. in *Machine Learning for Brain Disorders* **197**, 511–532 (2023).
172. Yang, Z. *et al.* A deep learning framework identifies dimensional representations of Alzheimer's Disease from brain structure. *Nat. Commun.* **12**, 7065 (2021).
173. Budd Haeberlein, S. *et al.* Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease. *J. Prev. Alzheimer's Dis.* **9**, 197–210 (2022).
174. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *N. Engl. J. Med.* 9–21 (2022). doi:10.1056/NEJMoa2212948
175. Castro, D. C., Walker, I. & Glocker, B. Causality matters in medical imaging. *Nat. Commun.* **11**, 1–10 (2020).
176. Scelsi, M. A. *et al.* Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci. *Brain* **141**, 2167–2180 (2018).
177. Jones, D. *et al.* A computational model of neurodegeneration in Alzheimer's disease. *Nat. Commun.* **13**, (2022).
178. Verdi, S., Marquand, A. F., Schott, J. M. & Cole, J. H. Beyond the average patient: How neuroimaging models can address heterogeneity in dementia. *Brain* **144**, 2946–2953 (2021).
179. Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* **163**, 115–124 (2017).
180. Cole, J. H., Marioni, R. E., Harris, S. E. & Deary, I. J. Brain age and other bodily 'ages': implications for neuropsychiatry. *Mol. Psychiatry* **24**, 266–281 (2019).

181. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
182. Zhang, X. *et al.* Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc. Natl. Acad. Sci.* E6535–E6544 (2016). doi:10.1073/pnas.1611073113
183. Edison, P. *et al.* Amyloid, hypometabolism, and cognition in Alzheimer disease: an [11C]PIB and [18F]FDG PET study. *Neurology* **68**, 501–8 (2007).
184. Butterfield, D. A. & Halliwell, B. Oxidative stress , dysfunctional glucose metabolism and Alzheimer disease. *Nat. Rev. Neurosci.* **20**, (2019).
185. Wiseman, F. K. *et al.* A genetic cause of Alzheimer disease: Mechanistic insights from Down syndrome. *Nature Reviews Neuroscience* **16**, 564–574 (2015).
186. Iadecola, C. The overlap between neurodegenerative and vascular factors in the pathogenesis of dementia. *Acta Neuropathologica* **120**, 287–296 (2010).
187. Iadecola, C. Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nat. Rev. Neurosci.* **5**, 347–360 (2004).
188. Bell, R. D. & Zlokovic, B. V. Neurovascular mechanisms and blood-brain barrier disorder in Alzheimer's disease. *Acta Neuropathologica* **118**, 103–113 (2009).
189. Vogels, T., Murgoci, A. N. & Hromádka, T. Intersection of pathological tau and microglia at the synapse. *Acta neuropathologica communications* **7**, 109 (2019).
190. Terada, T. *et al.* In vivo direct relation of tau pathology with neuroinflammation in early Alzheimer's disease. *J. Neurol.* **266**, 2186–2196 (2019).
191. Heneka, M. T. *et al.* Neuroinflammation in Alzheimer's disease. *The Lancet Neurology* **14**, 388–405 (2015).
192. Eshaghi, A. *et al.* Progression of regional grey matter atrophy in multiple sclerosis. *Brain* (2018). doi:10.1093/brain/awy088
193. Wojcik, C. *et al.* Staging and stratifying cognitive dysfunction in multiple sclerosis. *Mult. Scler. J.* **28**, 463–471 (2022).
194. *et al.* Neuroimaging biomarkers define neurophysiological subtypes with distinct trajectories in schizophrenia. *Nat. Ment. Heal.* **1**, 186–199 (2023).
195. Wen, J. *et al.* Characterizing Heterogeneity in Neuroimaging, Cognition, Clinical Symptoms, and Genetics among Patients with Late-Life Depression. *JAMA Psychiatry* **79**, 464–474 (2022).
196. Li, M. *et al.* Identifying the Phenotypic and Temporal Heterogeneity of Knee Osteoarthritis: Data From the Osteoarthritis Initiative. *Front. Public Heal.* **9**, 1–10 (2021).
197. Zhou, Y. *et al.* A foundation model for generalizable disease detection from retinal images. *Nature* (2023). doi:10.1038/s41586-023-06555-x
198. Young, A. L. *et al.* Disease progression modeling in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* **201**, 294–302 (2020).
199. Aglinskas, A., Hartshorne, J. K. & Anzellotti, S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science (80-.).* **376**, 1070–1074 (2022).
200. Archetti, D. *et al.* Inter-cohort validation of sustain model for alzheimer's disease. *Front. Big Data* **4**, 1–13 (2021).
201. Archetti, D. *et al.* Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. *NeuroImage Clin.* **24**, (2019).
202. Golriz Khatami, S., Salimi, Y., Hofmann-Apitius, M., Oxtoby, N. P. & Birkenbihl, C. Comparison and aggregation of event sequences across ten cohorts to describe the consensus biomarker evolution in Alzheimer's disease. *Alzheimer's Res. Ther.* **14**, 1–14 (2022).
203. Salimi, Y., Domingo-Fernández, D., Bobis-Álvarez, C., Hofmann-Apitius, M. & Birkenbihl, C. ADataViewer: exploring semantically harmonized Alzheimer's disease cohort datasets. *Alzheimer's Res. Ther.* **14**, 1–12 (2022).
204. Young, A. L., Oxtoby, N. P., Ourselin, S., Schott, J. M. & Alexander, D. C. A

simulation system for biomarker evolution in neurodegenerative disease. *Med. Image Anal.* **26**, (2015).
205. Dadgar-kiani, E. *et al.* Article Mesoscale connections and gene expression empower whole-brain modeling of a -synuclein spread , aggregation , and decay dynamics ll ll Mesoscale connections and gene expression empower whole-brain modeling of a -synuclein spread , aggregation , an. *CellReports* **41**, 111631 (2022).
206. Marinescu, R. V. *et al.* TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. *arXiv* (2018).
207. Marinescu, R. V *et al.* The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge: Results after 1 Year Follow-up. *J. Mach. Learn. Biomed. Imaging* **19**, 1–60 (2021).
208. Bron, E. E. *et al.* Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *Neuroimage* **111**, 562–579 (2015).
209. Bron, E. E. *et al.* Ten years of image analysis and machine learning competitions in dementia. *Neuroimage* **253**, 119083 (2022).
210. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* **14**, 535–562 (2018).

Figures

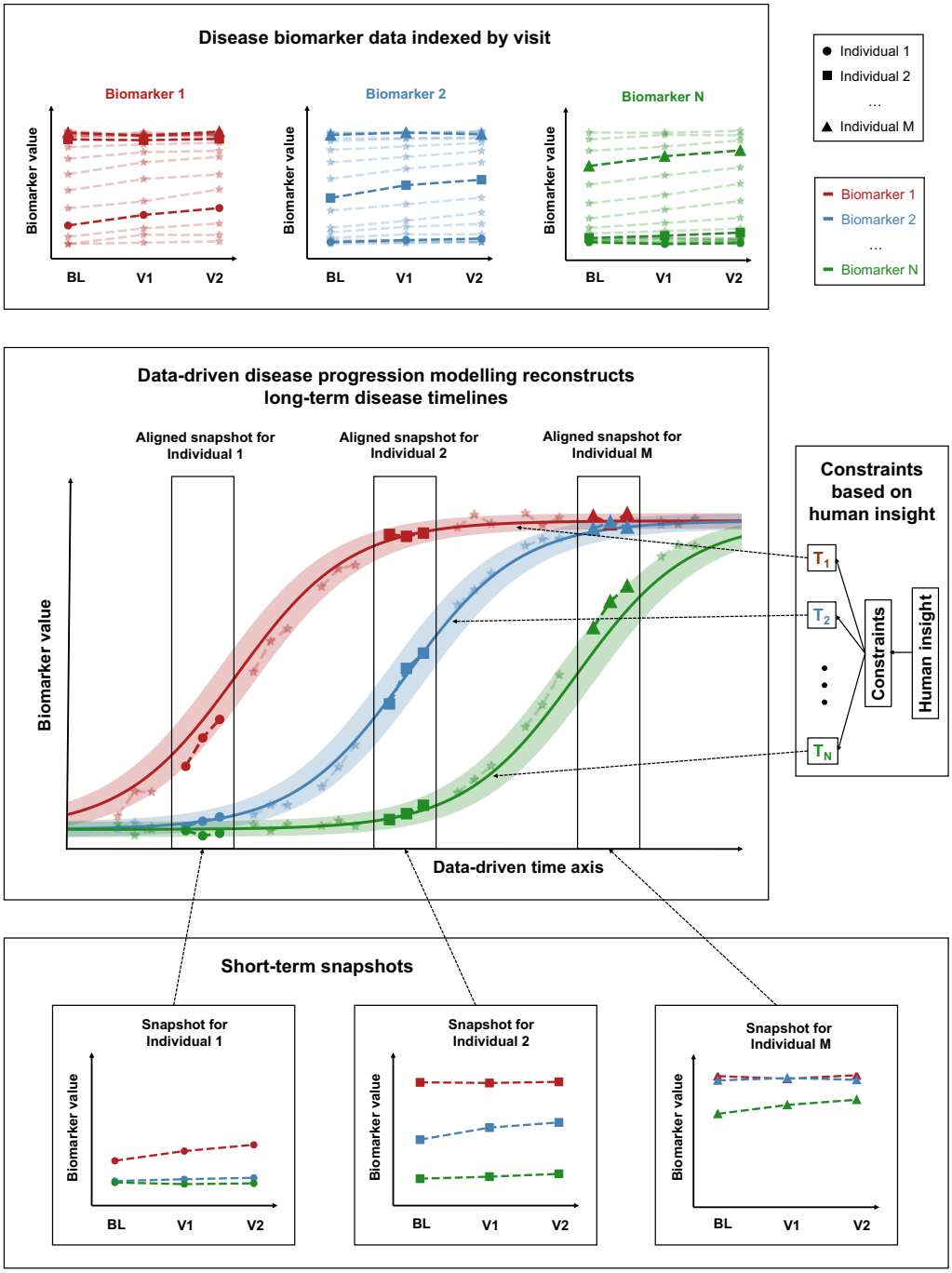

Figure 1. Framework for data-driven disease progression modelling. Typical study datasets (top) consist of heterogeneous short-term biomarker trajectories indexed by study visit (BL: baseline, V1: visit 1, V2: visit 2). Temporal heterogeneity (differences in the underling disease stages of individuals at baseline) is a key unresolved contributor to the observed heterogeneity in disease biomarker timelines. Data-driven disease progression models (middle left) disentangle temporal heterogeneity by temporally aligning short-term snapshots (bottom) to reconstruct long-term trajectories (red, blue, and green trajectories) subject to a set of constraints based on human insight (middle right). Disease progression models (middle left) simultaneously learn a data-driven time axis (x axis), a set of biomarker trajectories (red, blue, and green trajectories) and the alignment of snapshots (boxes).

Figure 2. Collage of selected phenomenological Data-Driven Disease Progression Models trained on observed data from studies of Alzheimer's disease, selected to illustrate the various model types from discrete (top-left) to continuous (top-right) to spatiotemporal (lower left) and subtyping (lower right). Top left: Event Based Model of Young et al.[38] showing the most likely ordering (vertical axis) and uncertainty (horizontal axis) of a multi-modal set of biomarkers in Alzheimer's disease. Top right: Latent Time Joint Mixed Model of Li et al.[62] showing continuous trajectories for a similar set of biomarkers. Lower left: Alzheimer's Disease Course Map of Koval et al.[70] showing estimated trajectories of cognitive decline, brain atrophy, brain shape changes, and hypometabolism in Alzheimer's disease. Lower right: Subtype and Stage Inference algorithm of Young et al.[40] showing three subtypes of Alzheimer's disease with distinct timelines of regional brain atrophy.

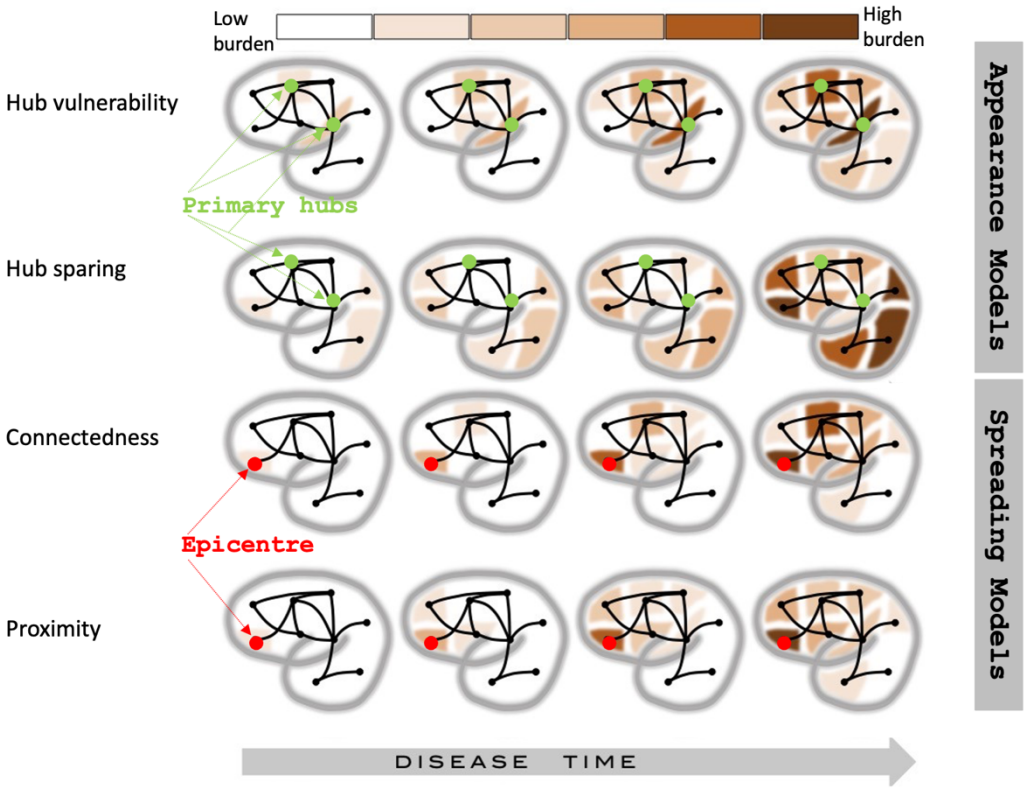

Figure 3. Simple network illustration of hypotheses that underpin pathophysiological models. Each row/model reflects a distinct hypothesis on how pathology appears (top two rows) or spreads (bottom two rows) in the brain, emulating hypothetical mechanisms. In network models the pathology burden is proportional to regional network properties. The first and second rows reflect mechanisms based on connectedness of regions (brain connectivity network is shown in black), without any notion of physical spreading. In the first row, the more connected, or “hub”-like a region, the more quickly it accumulates pathology; this reflects a mechanism in which “wear and tear” makes them vulnerable. The second row shows the opposite: the more isolated a region, the more quickly pathology accumulates, reflecting a “use it or lose it” mechanism. In the bottom two rows, pathology appears first in one “epicentre”. In the third row, pathology burden is proportional to how directly connected each region is to that epicentre, reflecting spreading through brain connectivity. Thus pathology appears soonest in regions most directly connected to the epicentre. In the fourth row, pathology burden is proportional to proximity to the epicentre, reflecting a mechanism that does not use the brain connectome, so the most proximal regions accumulate pathology most quickly.

Figure 4. Examples of pathophysiological models trained on observed data from studies of neurodegenerative diseases. (a) Predicted group atrophy patterns of Alzheimer's disease patients at different time-points using the network diffusion model presented in Raj et al.[116,119] which enforces "prion-like" spread; (b) Observed vs predicted patterns of tau deposition at different time-points from Vogel et al.[125] using the epidemic spreading model of Iturria-Medina et al.[118] (c) Top: MR images of a "typical" Alzheimer's disease patient; middle: simulated toxin protein evolution based on an initial seeding in the brain stem using the propagation model of Weickenmeier et al.[121]; bottom: simulated toxic protein-induced atrophy across the coronal slice, again using the model presented in Weickenmeier et al.[121]. (d) Disease progression patterns vs predicted patterns of atrophy severity at different time-points using the topological profile model of Garbarino et al.[94], which identifies a characteristic combination of the main mechanisms related to appearance and spread for Alzheimer's subjects.

Figure 5. Example applications of Data-Driven Disease Progression Modelling demonstrating the diversity of insight produced for applications such as providing biological insight, disease subtyping, and temporal stratification. Clockwise from top-left: **A.** Villemagne et al.[52] provide biological insight into the timeline of amyloid accumulation in Alzheimer's disease using a differential equaion model; **B.** Vogel et al.[155] demonstrate that subtype assignments from the Subtype and Stage Inference algorithm[40] predict cognitive decline in Alzheimer's disease; **C.** Young et al. [38] (left) and Oxtoby et al.[164] (right) demonstrate that temporal stratification into data-driven stages using an event-based model in Alzheimer's disease separates diagnostic groups (left) and predicts treatment response (right).

Tables

<table>
<tr><th></th><th colspan="4">Phenomenological</th><th colspan="3">Pathophysiological</th></tr>
<tr><th></th><th>Discrete</th><th>Continuous</th><th>Spatiotemporal</th><th>Subtyping</th><th>Network</th><th>Dynamical systems</th><th>Mechanistic combinations</th></tr>
<tr><td>Utility</td><td colspan="2">Data-driven disease patterns.
Temporal stratification.
Progression prognosis.</td><td>+ Enhanced spatial localisation of disease changes.</td><td>+ Subtype stratification.</td><td>Evaluating competing mechanistic hypotheses.</td><td>+ Pathophysiological parameter estimates.</td><td>+ Combinations and possibly interactions.</td></tr>
<tr><td>Input</td><td colspan="2">Scalar (e.g. spreadsheet).</td><td>Imaging data.</td><td rowspan="3">Inherited from chosen phenomenological model.</td><td>Brain maps of connectivity.</td><td>Brain maps, Imaging.</td><td rowspan="2">Inherited from chosen pathophysiological model.</td></tr>
<tr><td>Output</td><td>Sequence of states.</td><td colspan="2">Trajectories.</td><td>Expected sequence of pathology progression for evaluation against imaging changes.</td><td>Trajectories, Pathophysiological parameters.</td></tr>
<tr><td>Constraint Level</td><td colspan="2">Low.
e.g., monotonic</td><td>Moderate.
+ spatial constraints</td><td>Very High.
Connectivity entirely defines pathology pattern.</td><td>Relatively High. Only estimate a few key pathophysiological parameters.</td><td>Moderate. Varies according to the number of mechanisms and pathophysiological parameters for each mechanism.</td></tr>
</table>

Table 1. Table comparing the typical attributes of current phenomenological and pathophysiological models.