Simulating the multicausality of Alzheimer's disease with system dynamics

Uleman, J.F.; Melis, R.J.F.; Ntanasi, E.; Scarmeas, N.; Hoekstra, A.G.; Quax, R.; Olde Rikkert, M.G.M.; Alzheimer's Disease Neuroimaging Initiative

DOI
10.1002/alz.12923

Publication date
2023

Document Version
Final published version

Published in
Alzheimer's and Dementia

Received: 28 July 2022 | Revised: 25 November 2022 | Accepted: 15 December 2022

DOI: 10.1002/alz.12923

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

ALTERNATE FORMAT RESEARCH ARTICLE

Simulating the multicausality of Alzheimer's disease with system dynamics

Jeroen F. Uleman[1,2] | René J. F. Melis[2,3] | Eva Ntanasi[4] | Nikolaos Scarmeas[4,5] | Alfons G. Hoekstra[6] | Rick Quax[2,6] | Marcel G. M. Olde Rikkert[1] | the Alzheimer's Disease Neuroimaging Initiative[#]

[1]Department of Geriatric Medicine, Radboudumc Alzheimer Center, Donders Institute for Medical Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands

[2]Institute for Advanced Study, University of Amsterdam, Amsterdam, The Netherlands

[3]Department of Geriatric Medicine, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands

[4]Department of Neurology, Aiginition Hospital, National and Kapodistrian University of Athens Medical School, Athens, Greece

[5]Department of Neurology, The Gertrude H. Sergievsky Center, Taub Institute for Research in Alzheimer's Disease and the Aging Brain, Columbia University, New York, USA

[6]Computational Science Lab, Faculty of Science, Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

Correspondence
Jeroen Uleman, Department of Geriatric Medicine, Radboudumc Alzheimer Center, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Geert Grooteplein Zuid 10, 6500HB, Nijmegen, The Netherlands.
Email: jeroen.uleman@radboudumc.nl

Funding information
American Alzheimer's Association, Grant/Award Number: GEENA-Q-19-595225; NWO project Social HealthGames, Grant/Award Number: 645.003.002; Netherlands Organization for Health Research and Development, Grant/Award Number: 531003015; Alzheimer's Association, Grant/Award Number: IIRG-09-133014; National Institutes of Health, Grant/Award Number: U01 AG024904

Abstract

Introduction: In Alzheimer's disease (AD), cognitive decline is driven by various interlinking causal factors. Systems thinking could help elucidate this multicausality and identify opportune intervention targets.

Methods: We developed a system dynamics model (SDM) of sporadic AD with 33 factors and 148 causal links calibrated with empirical data from two studies. We tested the SDM's validity by ranking intervention outcomes on 15 modifiable risk factors to two sets of 44 and 9 validation statements based on meta-analyses of observational data and randomized controlled trials, respectively.

Results: The SDM answered 77% and 78% of the validation statements correctly. Sleep quality and depressive symptoms yielded the largest effects on cognitive decline with which they were connected through strong reinforcing feedback loops, including via phosphorylated tau burden.

Discussion: SDMs can be constructed and validated to simulate interventions and gain insight into the relative contribution of mechanistic pathways.

KEYWORDS
Alzheimer's disease (AD), complex system, complexity, intervention, modifiable risk factors, multicausal, multifactorial, multiscale, prevention, protective factor, risk factor, system dynamics, systems thinking

[#]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found on ADNI's website.

RESEARCH IN CONTEXT

1. **Systematic review**: We reviewed Alzheimer's disease (AD) studies related to complexity science. We observed an increasing interest in multicausal AD paradigms and applying systems thinking methods to AD. To our knowledge, the system dynamics model (SDM) we present is the most comprehensive computational model of AD's multicausality to date.
2. **Interpretation**: We demonstrated that SDMs can be systematically constructed, estimated, and validated based on literature, expert knowledge, and empirical data. As a use-case, we simulated interventions to rank modifiable risk factors for cognitive decline prevention. This ranking, in which sleep quality and depressive symptoms had the largest effects, was in line with observational (77%) and interventional (78%) meta-analyses.
3. **Future directions**: Our SDM offers a basis for further elucidating AD's multifactorial etiology using additional scientific knowledge sources. This would require an iterative cycle in which neuroscience researchers and computational scientists share their resources in a joint approach.

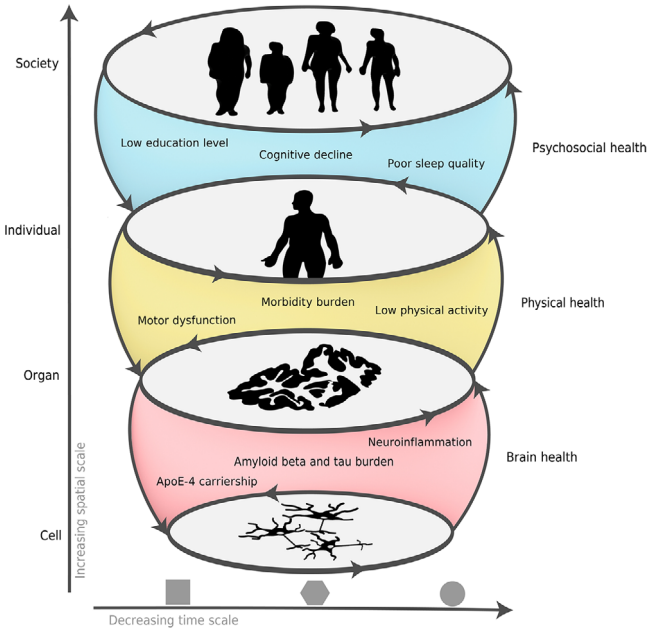

FIGURE 1 The multiscale multicausality paradigm of sporadic Alzheimer's disease. Decreasing time scales (x-axis) are shown with examples of constant or slowly changing variables on the left (square), for example, low education level, gradually changing in the middle (hexagon), for example, cognitive decline, and rapidly changing shown on the right (circle), for example, neuroinflammation. On the y-axis, relevant spatial scales are shown in increasing order. Across these spatiotemporal scales, opportunities can be identified for promoting brain health (e.g., pharmacological interventions), physical health (e.g., lifestyle interventions) and psychosocial health (e.g., psychological interventions)[4]

1 | NARRATIVE

1.1 | Contextual background

Alzheimer's disease (AD) is increasingly regarded as a complex disorder driven by a multitude of causal factors.[1–3] This multicausality includes neurobiological and psychosocial mechanisms that underlie the loss of neuronal functioning from which AD's cognitive and behavioral features emerge.[4] To better understand this complex process of cognitive decline, a growing number of researchers in the field call for the adoption of complexity science methods and, in particular, a systems thinking approach.[5–8] Complexity science in this field generally means research focused on understanding how mechanisms or processes and their connecting pathways lead to emergent phenomena such as resilience and cognitive decline.[8] Systems thinking is a related viewpoint that moves beyond monocausal characterizations by considering the multifactorial etiology of complex disorders as a whole and examining the underlying relationships across space and time.[9]

Through the lens of systems thinking, the multiscale multicausality paradigm of AD (Figure 1) may serve as a unifying framework to help connect the dots of AD's putative upstream variables (e.g., vascular[2,3] and metabolic factors,[10] inflammation,[11] comorbid conditions and other risks[12]) and downstream neurodegenerative processes (e.g., massive loss of synaptic and neuronal functioning, and cognitive decline). These variables include the pathophysiological processes in Figure 1: interactions among apoE-4 carriership, neuroinflammation,[11] and the accumulation of amyloid and phosphorylated tau, resulting in dysfunction of neurons and neuronal networks[13,14] at the cell and tissue scales, disrupting brain health. On the other side of the spatial scale spectrum, low education level[12] can result in less physical activity[15] and comorbidities, such as type 2 diabetes, hypertension, dyslipidemia, and obesity,[16] disrupting psychosocial and physical health at the individual organism scale. Acting across these spatial scales, such factors can influence one another and, ultimately, AD's emergent multi-system organ failure and cognitive disorder through a complex web of causal links.[4] This multicausality is governed by a large number of feedback loops, including cross-scale loops that are separated in space or time and can thus be difficult to recognize in research that focuses on specific scientific disciplines.[17] This involvement of feedback loops also suggests that there is no set order of events as downstream variables also influence upstream variables, such as cognitive decline leading to low physical activity,[18] limiting its protective effect against, for instance, chronic inflammation, vascular dysfunction, and other AD risk factors.[19] To characterize and analyze the combined effect of such processes, they could first be mapped out and then studied in silico to understand potential mechanistic routes. In turn, these routes can be validated empirically.

Uncovering and subsequently integrating the web of multicausality that drives AD into a single computational model is equally urgent as it is challenging.[20] Namely, comprehensive computational

models are needed to connect current theories and existing data and to test and validate putative causal links. Such models can also assess the relative importance of different risk factors because, unlike non-mechanistic models that can only predict outcomes (e.g., for risk assessment), implementing causal mechanisms allows for simulating interventions.[21,22] However, developing such comprehensive models will require iterative development, many existing AD data sets and extensive multidisciplinary collaboration.

As the starting point for developing a computational model,[20] we previously developed a conceptual multicausal mapping in the form of a causal loop diagram (CLD), which summarized expert-solicited causal links between many of the most important risk factors and pathophysiological mechanisms in sporadic AD.[4] This CLD resulted from the consensus of 15 domain experts through group model building[23] and was supported by a literature review. The CLD serves as a basis for system dynamics models (SDMs). When adequately validated, these SDMs can be used to scrutinize mechanistic pathways and explore interventional what-if scenarios,[20,24,25] for example, to design new randomized controlled trials (RCTs),[26,27] or explore interventions that are difficult, unethical, or even impossible to implement in RCTs.

Here, we describe the SDM we developed for sporadic AD based on this CLD.[4] We used empirical data from two AD studies to quantify the CLD's biopsychosocial pathways and feedback loops between AD pathophysiology, risk factors, and their effects on cognitive functioning. As a use-case, we then applied the SDM to AD prevention with the key objective of assessing the validity of the SDM's quantified causal structure by simulating interventions on 15 modifiable risk factors. Although dementia prevention is promising[26,28–31] and 40% and 56% of populations' risks are attributed to potentially modifiable risk factors,[12,32] interventions on these factors in RCTs have shown mixed results[33–38] with highly variable adherence.[30] This points to the prioritize the risk factors to be tested in clinical studies.[30] Insights obtained from an SDM can potentially accelerate this process.

1.2 | Study design

The SDM models a heterogeneous population of persons without dementia (at baseline) aged ≥ 55 years and simulates 5-year cognitive decline trajectories. The model was trained on data from both the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Hellenic Longitudinal Investigation of Aging and Diet (HELIAD) studies, which covered nearly all the risk factors and the pathophysiological mechanisms (together referred to as "nodes") included in the CLD.[4] The features available in these data sets were then used to operationalize each CLD node as closely as possible.

Once the data were selected, we started the SDM's development by labeling the nodes as auxiliaries (changing rapidly), stocks (changing gradually), and constants (very slowly or not changing). These labels were selected based on how fast we assumed the nodes to change relative to the assumed time scale of noticeable differences in cognitive decline (3 months) and within the total simulation time of 5 years. The SDM with corresponding labels is shown in Figure 2 and an interactive version is provided online.[39] Many of the nodes and links are aggregates that encompass more specific biological pathways, for example, neuronal dysfunction might encompass intracellular processes like mitochondrial dysfunction. If found to be important, such aggregates can be further elucidated in future modeling iterations.

We then estimated the strength of the 148 causal links in this SDM using data from the selected features from both studies. In our relatively short simulation time, these links were assumed to be linear (i.e., additive). We applied Bayesian inference to incorporate prior knowledge of the links' polarities (i.e., + or -) and characterized the parameters' uncertainty based on the corresponding posterior probability distribution. The parameter values with the highest posterior probability, that is, maximum a posteriori (MAP), and the parameters' standard deviations relative to the MAP are illustrated in Figure 2 as the arrow size and greyscale, respectively.

Once the model parameters were estimated, we tested the validity of the SDM. We first estimated its predictive accuracy by simulating 5-year trajectories of the 11 stocks for unseen individuals from independent test data. Specifically, we compared the MAP to a reference model in which the stocks changed over time with their population average in all individuals. Although point prediction is not the principal aim of SDMs,[40] this comparison assesses whether the SDM predicts individual trajectories better than making the naïve assumption that all individuals have equal progression.

Next, we simulated interventions on 15 modifiable risk factors by perturbing them with one standard deviation in 1000 randomly drawn samples from the posterior. We then ranked the risk factors by their effect on cognitive functioning (Alzheimer's Disease Assessment Scale [ADAS-cog-13 scale]) after 5 years and compared this ranking to two sets of validation statements[20,24] that were derived from meta-analyses. The first set contained 44 observational validation statements based on relative risk (RR) ratios from observational data.[12] For instance, one such statement asserted that an intervention on depressive symptoms (RR: 1.9 for dementia) should have a larger effect than an intervention on smoking (RR: 1.6). The second set contained nine interventional validation statements based on standardized mean differences (SMD) from RCTs and their durations.[29] For instance, one such statement asserts that an intervention on physical activity (SMD per year: 1.18) should have a larger effect than an intervention on obesity (SMD per year: 0.65).

1.3 | Main results

We assessed the predictive accuracy and ranking of simulated intervention effects to assess the SDM's validity. The average error on the test data was 8% lower for the MAP than the reference model (Section 3.7). The number of validation statements correctly answered by the SDM was 34/44 (77%) and 7/9 (78%) for the MAP on the observational and interventional statements, respectively, and significantly better than would be expected by chance ($p < 0.001$) (Section 3.7).

The modifiable risk factors with the strongest simulated intervention effects were depressive symptoms and sleep quality (Figure 3).

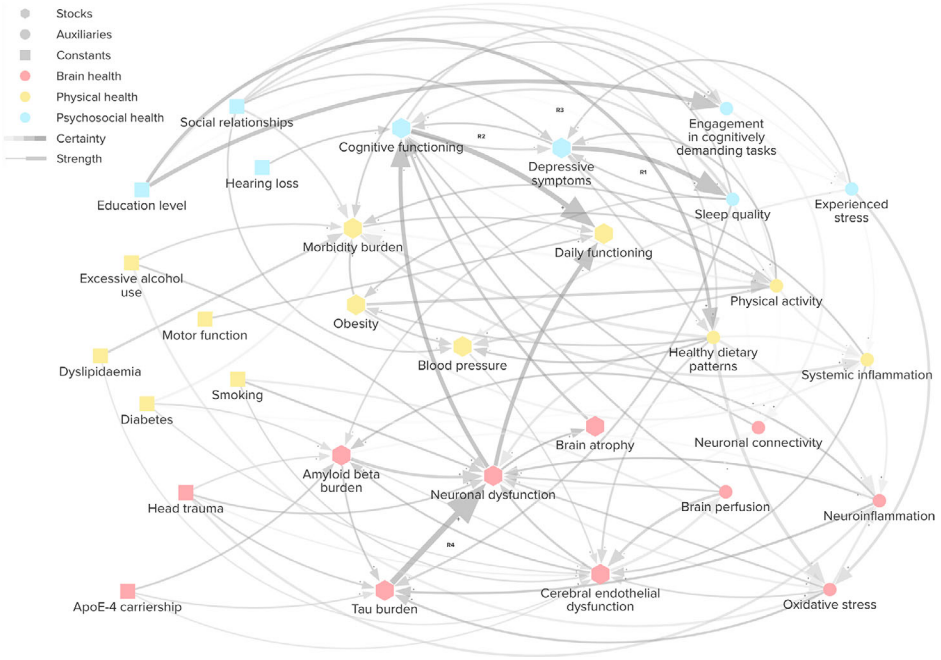

FIGURE 2 System dynamics model for AD with nodes from the causal loop diagram (CLD) as stocks, auxiliaries, or constants, and corresponding expert-solicited causal links (arrows) between them. R1-4 are examples of the many reinforcing feedback loops in the model (see Figure 4). Due to the CLD's focus on degenerative processes, the SDM does not contain any balancing loops.[4] The arrow size corresponds to the link's strength and is scaled to the standardized maximum a posteriori (MAP) parameter values. The grey scale corresponds to the uncertainty in the parameters, with darker meaning a smaller standard deviation relative to the MAP. Age was added as a predictor for the stocks and auxiliaries, and gender was added as a predictor for just the auxiliaries (not shown). An interactive version of this diagram can be found online[39]

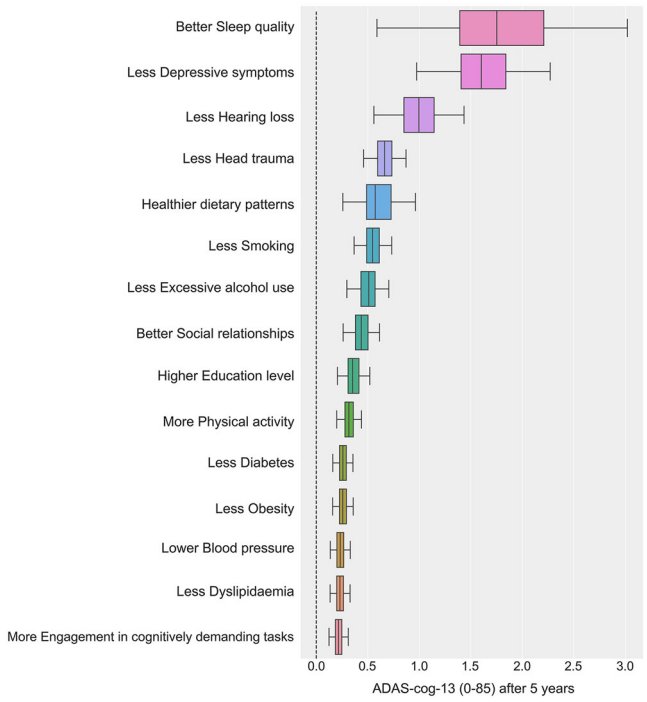

FIGURE 3 Effects of simulated interventions on the 15 modifiable risk factors that entailed the addition of +/- one standard deviation at baseline alternatively to each risk factor. The outcome (x-axis) was quantified as the improvement in cognitive functioning (i.e., reduction in ADAS-cog-13) after 5 years between simulations with and without intervention. The simulated interventions were conducted for 1000 posterior samples. The dotted line represents the absence of an effect

To demonstrate how SDMs can be used to gain insight into the pathophysiological processes underlying the effects of these risk factors, four potentially important feedback loops are shown in Figure 4. These loops are described below with estimated MAP parameters and corresponding 95% highest density intervals (i.e., the narrowest 95% credibility interval). The parameters can be interpreted as follows. For an auxiliary parameter, 0.5 means that one standard deviation change in X gives 0.5 change in Y. For a stock parameter (annotated by *), 0.5 means that one standard deviation change in X gives 0.5 change in Y over the simulation time of 5 years. First, both risk factors strongly interlink: the strongest effect on sleep quality came from depressive symptoms (0.30 [0.23, 0.36]), and sleep quality also influenced depressive symptoms (0.12* [0.06, 0.21]), forming a reinforcing feedback loop of length 2 (R1). Second, both factors have direct paths to cognitive functioning. Sleep quality had the third strongest effect on cognitive functioning (0.15* [0.06, 0.44]). The direct effect of depressive symptoms was less strong (0.12* [0.06, 0.19]), but cognitive functioning, in turn, had the strongest effect on depressive symptoms (0.15* [0.04, 0.28]), forming additional reinforcing feedback loops of length 2 and 3: R2 and R3, respectively. Finally, the strongest effect on cognitive functioning came from neuronal dysfunction (0.62* [0.43, 0.72]), on which the strongest effect came from (phosphorylated) tau burden (1.06* [0.91, 1.31]). Tau burden was affected by sleep quality (0.17* [0.06, 0.28]), forming a longer cross-scale feedback loop of length 5 (R4) that consists of multiple stocks, potentially playing an important mechanistic role in the long-term effect of poor sleep quality on cognitive decline. Other pathways from sleep quality and depressive symptoms,

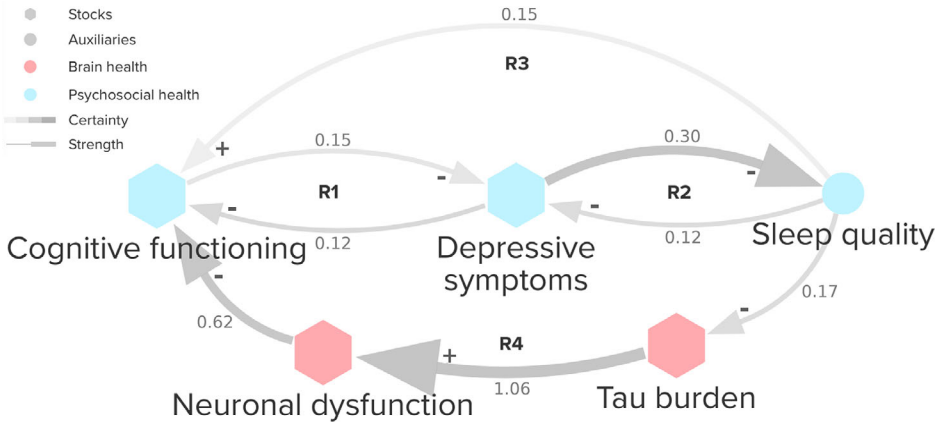

FIGURE 4 Interlinking feedback loops of the two highest ranked modifiable risk factors (see Figure 3): sleep quality and depressive symptoms. The arrow size corresponds to the link's strength and is scaled to the standardized maximum a posteriori (MAP) parameter values. The grey scale corresponds to the uncertainty in the parameters, with darker meaning a smaller standard deviation relative to the MAP

e.g., via experienced stress (0.04 [0, 0.11], 0.03 [0, 0.07], respectively; not shown in Figure 4), appear to be less important.

1.4 | Study conclusions and implications

We demonstrated that empirical data can be combined with expert-solicited knowledge regarding AD's system-wide multicausality in a single computational model to improve our understanding of the mechanisms underlying cognitive decline. Furthermore, incorporating additional expert-solicited knowledge regarding the parameters' polarities with Bayesian inference improved the simulated interventions while only minimally impacting the predictive accuracy (Section 3.7). While our model is intended as a proof of concept and requires further refinement, it already simulated validated risk factor interventions based on its inputs from one expert group,[4] two data sets (ADNI and HELIAD), and assuming linear functional relationships. Therefore, we conclude that system dynamics is an effective methodology for synthesizing knowledge on complex mechanisms that underlie AD and for quantifying intervention effects on a broad range of risk factors. We expect this may also be the case for other complex disorders.

The risk factors with the strongest simulated intervention effects were sleep quality and depressive symptoms (Figure 3). Interestingly, although they strongly correlate with AD,[41,42] these factors can also be symptoms of the disorder, which complicates teasing out cause and effect in observational studies.[12] Nevertheless, in the SDM, their effects on cognitive functioning are likely not the result of reverse causality. This is because their effects on cognitive functioning are modeled over time, the reverse effects from cognitive functioning are also modeled, and strong mechanistic pathways can be identified that may underlie their effects (e.g., the reinforcing feedback loops in Figure 4). Consequently, we hypothesize that improving sleep quality and limiting depressive symptoms in multidomain RCTs might improve the efficacy of the interventions. However, evidence on the effectiveness of preventive treatments on these two factors is insufficient.[29,43,44] Thus far, neither factor has been directly used as an intervention in multidomain RCTs[30] but such RCTs *are* currently being conducted.[45,46] The results of these trials will be another test for the validity of the SDM.

1.5 | Limitations and future directions

As all models are inevitably incomplete,[47] the SDM presented here also requires further elaboration. For instance, although the CLD is based on the consensus of 15 experts and a supporting literature review, important causal links may have been missed or falsely added. Hence, future work might entail the development of a comprehensive method for evidence triangulation[48] that combines evidence for the causal links from group model building,[23] scientific literature, and data-driven causal discovery methods.[22] That said, the possible bias introduced in our results by small mistakes in the SDM's struture (e.g., literature bias) is likely small as, for instance, the correlation between the simulated intervention ranking and the number of outgoing links of the risk factors is not statistically significant (Spearman's $r = 0.40$, $p = 0.14$). Important nodes may also be missing in the SDM. One reason is that such nodes, for example, air pollution,[12] were not part of the CLD. Another reason is that certain CLD nodes were not available in the data. For instance, our data contained insufficient information on glymphatic system functioning, which led us to replace it with a direct causal link (Section 3.2). However, the glymphatic system may be an important mechanism that not only relates sleep quality to amyloid beta and tau burden,[4,49] but head trauma (another high-ranked factor, Figure 3) as well.[4] Including more intermediary variables like the glymphatic system in the SDM could contribute to a deeper mechanistic understanding of AD risk factors. Future work could then also entail simulating interventions on such pathophysiological mechanisms through pharmacological interventions.

Another related limitation is the linear approximation of the short-term causal effect of each link. Although we found this assumption valid within the relatively short simulation time (section 3.2), it prevented us from simulating longer than 5 years. Longer simulation times

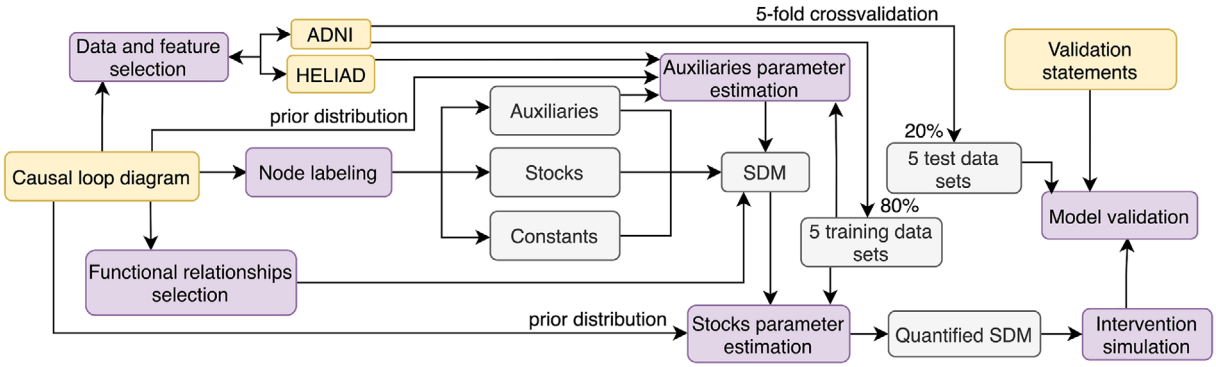

FIGURE 5 Schematic overview of the procedure we utilized for obtaining the system dynamics model from the causal loop diagram for AD.[4] Yellow: model inputs; purple: general modeling steps

will be ultimately required to study AD prevention from a life-course perspective, which will be needed to identify optimal risk factor interventions for different age groups.[26,30] To that end, future work may involve group model building sessions that target the functional forms of causal links,[20] combined with a model selection procedure that identifies nonlinear links from the data. Such procedures would likely improve the predictive accuracy of the SDM, which, while 8% better than predicting only population averages (i.e., the reference model, section 3.6), was still relatively low, with the simulated trajectories not generally matching individual deviations from the stock's averages well (section 3.7). However, a data-driven model selection procedure would require large, high-quality data sets containing thousands of subjects and features covering most or all of the nodes in the SDM, highlighting the need for harmonization efforts of longitudinal AD data sets.

Given the above limitations, we reiterate that the SDM is not a final and accurate representation of reality but a starting point to be systematically developed further. Here we applied such a systematic method, which can, in principle, be used to develop increasingly comprehensive SDMs that encompass the best-evidenced risk factors and pathophysiological processes and are predominantly limited by the current bounds of scientific knowledge. This limit is also an opportunity. By synthesizing scientific knowledge, SDMs can identify critical sources of uncertainty that should be further elucidated, for example, through the targeted collection of new data by neuroscience or related domain researchers. As such, we envision SDMs as part of an iterative inductive-deductive cycle, which takes various sources of scientific knowledge as input and produces identified knowledge gaps and falsifiable hypothesizes that can be tested in empirical research.

2 | CONSOLIDATED STUDY DESIGN

To develop the SDM, we followed an adaptation (Figure 5) of the systematic procedure for developing biopsychosocial SDMs from CLDs outlined in Crielaard et al.,[20] which is summarized in section 1.2 but outlined in further detail in sections 2 and 3.

2.1 | Data and feature selection

We used the Global Alzheimer's Association Interactive Network interrogator[50] to identify a small number of data sets that covered as many nodes from the CLD as possible. Ultimately, we selected ADNI, as it contains most pathophysiological variables and risk factors from the CLD and HELIAD, which covers several lifestyle factors missing in the ADNI data. Section 3 provides more details regarding the data processing (section 3.1.1) and selection of features to operationalize the nodes (section 3.1.4). Table 1 provides an overview of the selected features and their descriptive statistics at baseline.

2.2 | Node labeling

We made assumptions regarding the time scales of the CLD nodes to determine each node's label in the SDM as a stock, auxiliary or constant (Section 3.2.1). Stocks are each described by ordinary differential equations, which model the slope of the stock at a particular time given the values of the input nodes. A node is labeled as a stock if its time scale (i.e., the approximate time to reach a new stable value given a change in its input node values) is assumed to be within an order of magnitude of the assumed time scale of cognitive functioning (i.e., 3 months). In turn, auxiliaries are defined using (algebraic) regression equations, meaning that their value changes instantly to a new stable value given the values of their input nodes, as opposed to changing gradually according to a slope as a stock does. Finally, and on the opposite side of the time scale spectrum, constants do not change but may appear on the right-hand side of both the differential and algebraic equations. Section 3.2 provides further details on node labeling with sensitivity tests.

2.3 | Selecting functional relationships

To enable the identification of the SDM's differential (stock) and algebraic (auxiliary) equations, an assumption had to be made regarding the class of functional formulae to be considered. Given the considerable

TABLE 1 Baseline characteristics of the preprocessed ADNI ($N = 1762$) data features

CLD node	Data feature (units)	Mean	SD	Missing	Slope
	Diagnosis (0: cognitively normal, 1: mild cognitive impairment)	0.56	0.50	0%	
	Age (years)	73.0	6.98	0.2%	
	Gender (0: female, 1: male)	0.53	0.50	0.0%	
Auxiliaries					
Experienced stress	Cortisol plasma (ng/ml)	145	1.35	75%	
Systemic inflammation	TNF-alpha plasma (pg/ml)	6.76	1.86	75%	
Brain perfusion	Hippocampal CBF (ml/mg/min)	28.5	7.50	93%	
Oxidative stress*	Superoxide dismutase plasma (pg/ml)	45.7	1.74	75%	
Neuroinflammation	TNF-alpha CSF (pg/ml)	1.81	1.13	88%	
Neuronal connectivity*	Default mode network connectivity anterior-posterior ratio	0.99	0.21	93%	
Physical activity**	Physical activity score (1-10)	7.12	0.42	2%	
Healthy dietary patterns**	Mediterranean diet score (1-55)	34.9	1.01	1%	
Sleep quality**	Sleep Scale from the Medical Outcomes Study (1-54)	16.1	1.72	2 %	
Engagement in cognitively demanding tasks**	Cognitive activities score (0-12)	4.46	0.50	1%	
Stocks					
Cognitive functioning*	ADAS-cog-13 (0-85)	13.1	6.92	0.5%	0.96
Brain atrophy	Hippocampal volume (mm^3)	7132	1081	13%	−131
Neuronal dysfunction*	FDG-PET (g/ml)	1.27	0.13	34%	−0.02
Amyloid beta burden*	Amyloid beta CSF (pg/ml)	1053	455	45%	−19.9
Daily functioning*	Functional activities questionnaire (0-30)	1.84	3.45	0.7%	1.06
Morbidity burden	Combined morbidity score (0-1)	0.32	0.17	24%	0.02
Depressive symptoms	Geriatric depression scale (0-15)	1.30	1.38	0.1%	0.17
Obesity	Body mass index (kg/m^2)	27.2	4.95	0.8%	−0.14
Blood pressure	Pulse pressure (mmHg)	59.4	14.7	0.5%	−0.30
Tau burden	Phosphorylated tau CSF (pg/ml)	25.5	13.1	45%	0.41
Cerebral endothelial dysfunction	White matter hyperintensity volume (cm^3)	5.67	8.85	35%	0.33
Constants					
Diabetes	Diabetes (yes/no)	0.15	0.35	24%	
Dyslipidemia	Dyslipidemia (yes/no)	0.73	0.44	24%	
Social relationships	Currently married (yes/no)	0.75	0.43	0.4%	
Hearing loss	Hearing impairment (yes/no)	0.10	0.30	0%	
Smoking	History of smoking (yes/no)	0.40	0.49	24%	
Excessive alcohol use	History of alcohol abuse (yes/no)	0.04	0.19	24%	
Head trauma	Traumatic brain injury (yes/no)	0.05	0.22	24%	
ApoE-4 carriership	ApoE-4 alleles (0-2)	0.49	0.63	1.9%	
Education level	Received education (years)	16.2	2.67	0%	
Motor function*	Motor strength impairment (yes/no)	0.03	0.17	0%	

For stocks, the average slopes per year over all the individuals over the first 5 years in the data are also shown.
Abbreviations: CSF, cerebrospinal fluid; CBF, cerebral blood flow.
* These nodes have the opposite interpretation in the CLD as in the data (e.g., ADAS-cog-13 measures cognitive dysfunction, hence its slope is positive rather than negative).
**These nodes are imputed into ADNI from the HELIAD data; the reported characteristics are from the imputed data. The corresponding characteristics in the HELIAD ($N = 1924$) data are provided in section 3.1.1.

complexity in terms of model parameters (many of which represent time-dependent relationships) and feedback loops, and that nonlinear behavior can already arise from coupled first-order linear differential equations, we made the common assumption of local linear approximation. This assumption implies that the causal links are considered linear (i.e., additive) and first-order (i.e., only the first derivatives are used) in the relatively short simulation time we used (5 years) compared to the overall time frame of preclinical AD progression of several decades during which the individual links may become nonlinear. We verified this assumption in section 3.3. Section 3.4 provides the resulting model equations with the estimated model parameters.

2.4 | Parameter estimation

For estimation and validation purposes, we applied five-fold cross-validation and split the data five times so that each individual was both used to estimate the model parameters and assess the (out-of-sample) predictive accuracy of the SDM. After fitting a model without incorporating prior knowledge into the parameter estimates, we observed very high correlations between the parameters as well as unexpected polarities of parameter estimates with large uncertainties around them, suggesting that the parameters were underdetermined and could not be identified well from the data. This may be a consequence of the convoluted nature of the CLD, as reported by the group model building experts.[4] One way to mitigate this is by adding additional model assumptions,[21] so we next applied Bayesian inference to incorporate prior knowledge regarding the sign (+ or -) of the parameter estimates (which was already encoded in the links' polarities in the CLD). To this end, we used Markov Chain Monte Carlo methods to obtain a posterior probability distribution over the parameters of each training data fold based on both the data and prior knowledge. Section 3.5 provides further details on the parameter estimation.

2.5 | Model validation

We first conducted the behavior pattern test[20,40] to assess the (out-of-sample) predictive accuracy, i.e., generalizability, of the SDM. Since we conducted five-fold cross-validation, we simulated the individuals in each test set (20% of all individuals per fold) and compared the root mean squared errors (RMSEs) of the MAP (model with prior) and the model without prior to the RMSE of a reference model. This reference model assumed that all individuals increased with the same average slope from baseline for each stock. Finally, we compared the RMSEs in the test sets of each fold to the RMSE of the training sets, which suggested that the model did not over- or underfit the training data folds (section 3.7).

We then conducted the structure-oriented behavior test[20,40] to assess the plausibility of the simulated interventions in 1000 samples from the combined posterior over all data folds. We simulated the effect of interventions using a one-at-a-time sensitivity analysis in which we alternatively added (protective factors) or subtracted (risk factors) one standard deviation to the baseline values (stocks and constants) of 15 modifiable factors. In the case of an auxiliary risk factor, we added or subtracted a standard deviation as a constant to its equation. We then determined effect sizes for the risk factors by comparing the difference in cognitive functioning after a simulated trajectory of 5 years with and without an intervention. Finally, we ranked the risk factors by their effect sizes and compared this ranking to two sets of validation statements derived from previously published meta-analyses. For one set, based on observational data, we extracted 44 validation statements regarding 11 of the modifiable risk factors from the RR ratios reported by Livingston et al.[12] For the other set, based on RCTs, we extracted nine validation statements regarding four of the modifiable risk factors from the standardized mean differences and trial durations reported by the World Health Organization.[29] Section 3.7 provides further details on the model validation.

3 | DETAILED STUDY DESIGN AND RESULTS

3.1 | Data selection

3.1.1 | Data selection and preprocessing

ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD, aiming at AD detection at the earliest possible stage and supporting advances in AD prevention. Participants in ADNI were recruited at 57 memory clinic sites in the United States and Canada, with age ranges from 55 to 90 years and follow-up visits occurring every 6 months. The relevant ADNI data files were obtained from loni.adni.com, contained 2127 individuals, and were extracted and merged on subject ID and visitation code. We removed duplicate records (i.e., multiple measurements for the same subject ID and visitation code) and estimated relevant features (such as morbidity count, body mass index [BMI], and pulse pressure) from related variables. Either screening or baseline values were used as initial values, depending on their availability. If both were available for an individual, the baseline values were used. Given our specific interest in prevention, we excluded subjects who already had dementia at baseline from the data ($N = 365$). No other exclusion criteria were applied to remain representative of the highly heterogeneous population who could potentially develop late-onset (non-familial) AD, leaving 1762 individuals in the data.

Most features we selected (Table 1) were directly available in the ADNI data (such as ADAS-cog-13), but some features had to be extracted from the patients' medical histories. In particular, the combined morbidity score was calculated as the fraction of morbidities in the following categories: Psychiatric, Neurologic (other than AD), Cardiovascular, Respiratory, Hepatic, Endocrine-Metabolic, Gastrointestinal, Hematopoietic-Lymphatic, Renal-Genitourinary, and Malignancy. Diabetes was extracted by searching in the medical history (at baseline or month 6) for search terms: "diabetes", "diabetic", "insulin", and "DM", or having fasting glucose > 125 mg/dl at baseline. Dyslipidemia was similarly extracted using search terms:

"dyslipidemia", "hyperlipidemia", "hypercholesterolemia", "triglycerides", "cholesterol" and "lipids", or having a total cholesterol of > 200 mg/dl at baseline. Finally, head trauma was extracted by searching in medical history at baseline using search terms: "head injury", "TBI", "head trauma", "concussion", "fractured skull", and "skull fracture".

The number of individuals available for estimating the model parameters differed per procedure. The baseline means, standard deviations, and missing rates of the ADNI features are provided in Table 1. For the ADNI auxiliaries, the baseline data of plasma cortisol ($N = 434$), plasma TNF-alpha ($N = 220$), hippocampal cerebral blood flow ($N = 125$), plasma superoxide dismutase ($N = 223$), cerebrospinal fluid TNF-alpha ($N = 165$), and default-mode network connectivity ($N = 107$) of individuals were used. For the estimation of the stock parameters, we omitted individuals that did not have data available for all stocks and constants at baseline because a complete set of baseline values of the ADNI data was necessary to initialize the SDM simulations. This resulted in $N = 600$ individuals for the estimation of the stock parameters, for a total of 25051 data points over all individuals, stocks and time-points. These data ($N = 600$) were divided into 80% training data ($N = 480$) and 20% test data ($N = 120$) five times using five-fold cross-validation. We did not require that each of these $N = 600$ individuals had longitudinal data points available for all stocks. For the specific stocks, 98% (ADAS-cog-13), 99% (hippocampal volume), 51% (fluorodeoxyglucose-positron emission tomography [FDG-PET]), 99% (white matter hyperintensity [WMH] volume), 50% (amyloid beta), 98% (functional activities questionnaire), 17% (morbidity score), 98% (geriatric depression scale [GDS]), 100% (BMI), 100% (pulse pressure), and 50% (phosphorylated tau) of individuals had one or more longitudinal data points available. The average slopes of these stocks are provided in Table 1.

HELIAD is a population-based, multidisciplinary, collaborative study designed to estimate the prevalence and incidence of MCI and AD as well as other types of dementia and neuropsychiatric conditions of aging in the Greek population of 65 and older. In HELIAD, demographic, clinical and lifestyle factors were obtained at baseline and in the 3-year follow-up to examine the relationship between lifestyle factors and neuropsychiatric disease in older individuals. The HELIAD data set, which contained 2083 individuals, was selected from the GAAIN platform because it contains data on each of the lifestyle factors that were missing in the ADNI data. We also excluded subjects from HELIAD who had dementia at baseline ($N = 158$), leaving 1924 individuals.

For nodes that ADNI lacked data for, we selected from HELIAD the Sleep Scale from the Medical Outcomes Study (MOS-SS) for sleep quality, a score for physical activity, a score based on magazine/newspaper reading and museum visits for engagement in cognitively demanding tasks, and a Mediterranean diet score derived from a food frequency questionnaire for healthy dietary patterns. For these HELIAD variables, which were all auxiliaries, cognitive activity ($N = 803$; mean = 3.03, sd = 2.51), Mediterranean diet score ($N = 849$; mean = 33.4, sd = 4.53), physical activity score ($N = 492$; mean = 6.89, sd = 2.01), and the sleep scale ($N = 867$; mean = 17.7, sd = 7.68) were used for the parameter estimation. Since FDG-PET (neuronal dysfunction) and motor function were not available in HELIAD, the causal links (from the CLD) from neuronal dysfunction to sleep quality (replacing the effect of circadian misalignment in the CLD) and motor function to physical activity were omitted as predictor terms in the estimation of sleep quality and physical activity.

ADAS-cog-13 was unavailable as a predictor for physical and cognitive activity in the HELIAD data. Therefore, both were regressed using an ADAS-cog-13 score based on the Mini-Mental State Examination (MMSE) score (Pearson's r: -0.74) estimated from the ADNI data, resulting in the following conversion formula: $ADAS = 90.1 - 2.7 \times MMSE$.

3.1.2 | Data sets comparison

Generally, a trade-off exists between using more data sets (i.e., having greater coverage of the CLD nodes) and the compatibility of the parameters estimated from these data sets. Given this trade-off, we decided to use two data sets: ADNI and HELIAD. To support this decision, we compared the distributions of the predictors we used from both data sets (Figure 6) and the two confounding factors we adjusted for: age and gender. As can be seen, social relationships (marital status), cognitive functioning (MMSE), obesity (BMI), and depressive symptoms (GDS) were quite similar between the two data sets. The estimation of all auxiliaries was adjusted for age and gender. The feature that differed most between the two data sets, namely education level (years of followed education), was also adjusted in the estimation of the auxiliaries, except for sleep quality, as it was not included as a direct cause in the CLD.

3.1.3 | Sensitivity tests

To assess the impact of adding the HELIAD data, we conducted a sensitivity test in which we only used ADNI data to estimate the model (in which case sleep quality, healthy dietary patterns, physical activity, and engagement in cognitively demanding tasks were omitted from the model). This resulted in a very similar ranking of simulated intervention effects with depressive symptoms, head trauma, and hearing loss being the top three ranked risk factors, respectively. In addition, we conducted a sensitivity test in which we included only individuals who were cognitively normal (i.e., by excluding persons with mild cognitive impairment from the data). This likewise resulted in a very similar ranking with sleep quality, depressive symptoms, and hearing loss as the top three ranked risk factors, respectively.

3.1.4 | Feature selection

Based on their availability in the data sets, we selected features to operationalize the CLD nodes as closely in line as possible with the node definitions provided by the group model building experts during group model building.[4] In most cases, straightforward features were available, such as the number of years of received education

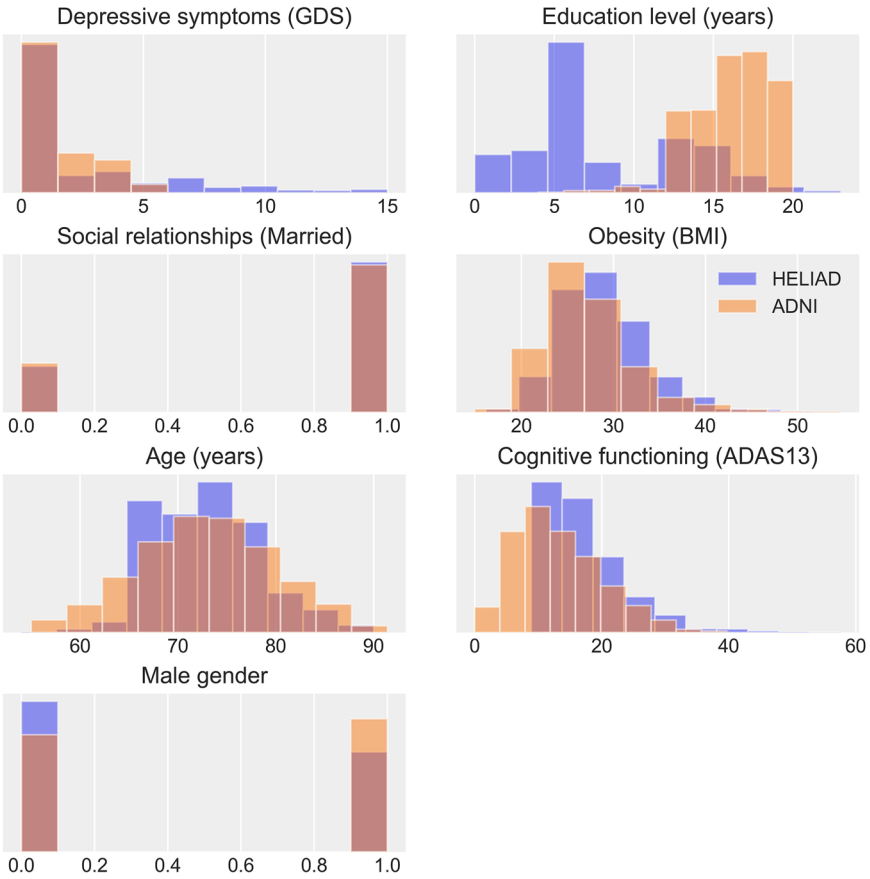

FIGURE 6 Histograms of features from the HELIAD and ADNI data sets. Education level is measured in years of received education, obesity is measured by the body mass index, social relationships is measured as marital status (1: currently married), cognitive functioning is measured by the Mini-Mental State Examination (in HELIAD), depressive symptoms is measured with the Geriatric Depression Scale, and age is measured in years

to operationalize education level. In other cases, only a proxy was available, such as marital status for social relationships. However, many different plasma and cerebrospinal fluid markers were available for systemic- and neuroinflammation. Combining these into a single measure using principal component analysis added no improvement in the model's accuracy compared to using single markers. Hence, we selected TNF-alpha as a single representative marker because it crosses the blood-brain barrier, both macrophages and microglia secrete it, and its inhibition may have therapeutic benefits.[51] As the primary variable of interest, we selected for cognitive functioning the cognitive subscale of the ADAS-cog-13 over alternatives like the MMSE or ADAS-cog-11 due to its greater responsiveness to AD progression, including potentially in subjects with predementia.[52]

3.2 | Node labeling

3.2.1 | Node labeling details

System dynamics was selected as an appropriate starting point for developing a computational model because it can already capture the CLD's complexity by describing the dynamics of various interrelated quantities over time in a single model. As a first approximation, it abstracts CLD nodes to three variable types (stocks, auxiliaries, constants) based on their assumed time scales. SDMs can also be used to identify which parts of the system require further elaboration, after which they can be flexibly coupled to more sophisticated types of models if needed,[20] for example, an agent-based model for simulating group interventions or a partial differential equation model for simulating spatially heterogeneous patterns of brain atrophy. Starting with more detailed models like these could introduce greater complexity at the level of individual nodes while potentially coming at the cost of not implementing the system of multicausality as a whole.

To label the nodes into stocks, auxiliaries, and constants, we followed the novel approach by Crielaard et al.[20] that utilizes assumed time scales of the CLD nodes (Figure 7). For example, obesity (BMI) is a stock as it would not drastically change from 1 week to another, even if its direct causes would. For the auxiliaries, we assumed that inside-body processes generally changed rapidly (e.g., inflammation, oxidative stress, and brain perfusion). A few exceptions were made to this rule based on the expert-solicited CLD definitions of the nodes. Specifically, we implemented as stocks: blood pressure, defined as "long term" in CLD,[4] amyloid beta and tau burden, defined by a gradual process of deposition,[53] neuronal dysfunction, involving neuronal damage and death, which is unlikely to occur on a short time scale, and cerebral endothelial dysfunction, operationalized as WMH volume, which accumulates over longer periods of time. However, physical activity, healthy dietary patterns, sleep quality, and engagement in cognitively demanding tasks were assumed to be auxiliaries as we consider them as states (rather than traits) that, relative to the scale of months, can change rapidly as a function of their inputs. For instance, more physical activity might take several weeks or months to alleviate depressive symptoms

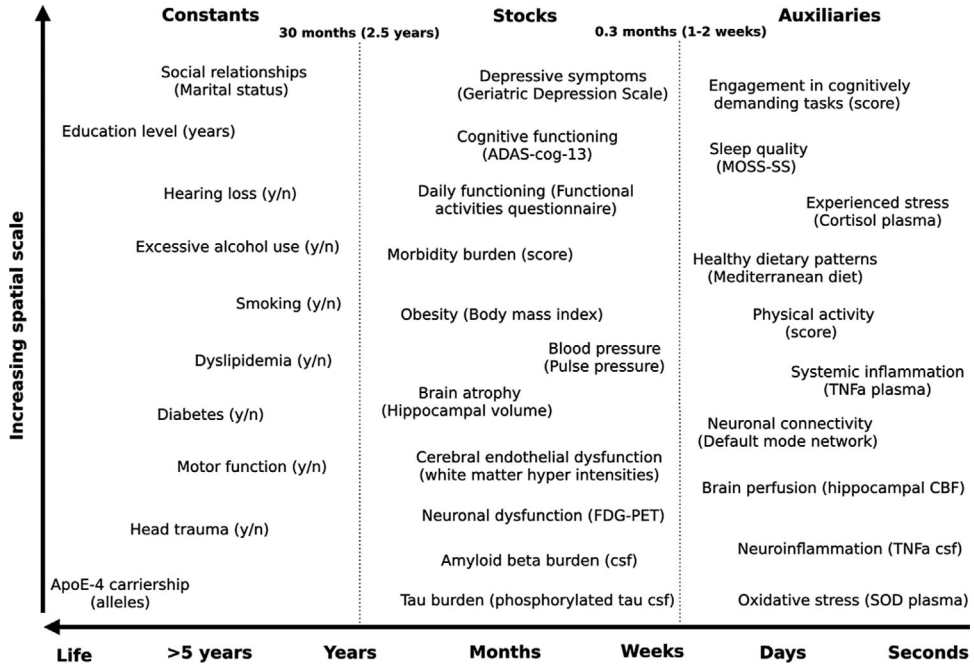

FIGURE 7 Scale-separation map with the assumed time scales of the nodes on the x-axis. Based on the assumed time scales (x-axis) relative to the assumed time scale of cognitive functioning (3 months), the nodes are divided into auxiliaries (< 1 to 2 weeks, corresponding to 0.3 months), stocks (within one order of magnitude of 3 months: 0.3 to 30 months), and constants (> 2.5 years, corresponding to 30 months)

(a stock), for example, by increasing brain-derived neurotrophic factor levels and modulating serotonergic neurotransmission,[15] whereas a decrease in depressive symptoms might improve a person's motivation for physical activity (an auxiliary) almost immediately. As constants, we selected the ("exogenous") nodes that do not have incoming links in the CLD, namely apoE-4 carriership, head trauma, and education level, as they have no way of changing their value in our model. Additionally, we checked how much the nodes changed over time. Several of them were operationalized with binary variables that changed in less than 8% of individuals over the utilized simulation time and were therefore assumed constant, in particular: smoking (< 8%), excessive alcohol use (< 1%), motor function (< 8%), diabetes (< 3%), dyslipidemia (< 2%), social relationships (< 1%), and hearing loss (< 8%). Except for diabetes and dyslipidemia (section 3.1), each of these constants was set to their baseline values because social relationships and excessive alcohol use changed in less than 1% of individuals, and smoking, motor function and hearing loss were almost completely missing over time (except for 6% to 8% of the individuals which were available but only after 4 years. For smoking, data were available after 1 year as well but at this point only 0.4% of people had changed their status).

3.2.2 | Sensitivity tests

We conducted three sensitivity tests to assess the impact of assumptions made in the labeling process. First, we conducted a sensitivity test in which we replaced diabetes and dyslipidemia (as constants) with fasting glucose and total cholesterol (as auxiliaries). Second, we conducted a sensitivity test in which we implemented depressive symptoms, obesity, and blood pressure as constants rather than stocks. Third, we conducted a sensitivity test in which we implemented depressive symptoms, tau burden, and blood pressure as auxiliaries rather than stocks. Each of these tests resulted in a very similar ranking, with sleep and depressive symptoms being the top two risk factors, hearing loss and head trauma being among the top five risk factors, together with either healthy dietary patterns or smoking. In addition, in each test, blood pressure, obesity, engagement in cognitively demanding tasks, diabetes, and dyslipidemia were among the lowest ranked risk factors.

3.2.3 | Replacing nodes that could not be operationalized

Not all nodes from the CLD were included in the SDM because they were insufficiently available in the utilized ADNI study data. These were wealth, microbleeds, lacunar infarcts, glymphatic system function, circadian misalignment, and damage to neurotransmitter systems. Fortunately, all these nodes, except for one, have either only a single incoming link from another node (single-in) or only one outgoing link to another node (single-out) or both (single-in single-out). In all these cases, the node can be effectively replaced by a direct link from the incoming to the outgoing node in a linear model.[54] The exception was glymphatic system function, which has two incoming (sleep quality and head trauma) and outgoing (amyloid beta and tau burden) links in

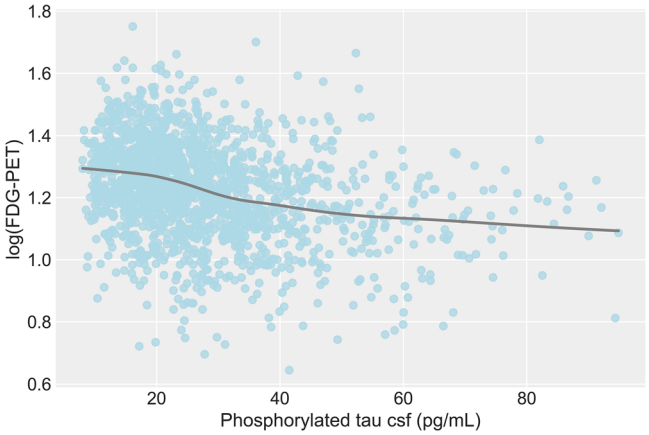

FIGURE 8 Scatter plot of the strongest causal link in the SDM that goes from tau burden (phosphorylated tau in the cerebrospinal fluid) to neuronal dysfunction (FDG-PET) with locally weighted scatterplot smoothing to identify trends in the data.

the CLD. However, since we operationalized head trauma as a binary constant that is only present in 5% of the simulated population, we reasoned that, for at least 95% of the individuals, the dynamics would be exclusively driven by sleep quality. Hence, glymphatic system function was replaced by direct links from sleep quality and head trauma to amyloid beta and tau burden. Although ADNI data were available for WMHs (a double-in, single-out node in the CLD), we used it as a marker to operationalize cerebral endothelial dysfunction.[55] This is reasonable since WMHs is a single-out node, and cerebral endothelial dysfunction is one of the two nodes with a causal link to WMHs, shares the same incoming links (brain perfusion) and has the same outgoing link (neuronal dysfunction) in the CLD.

3.3 | Validation of the local linear approximation assumption

We checked the validity of the local linear approximation assumption (i.e., that first-order linear differential equations can describe the causal links within the simulation time of 5 years) by log-transforming the stocks and predicting them with linear mixed models using either just Time, or Time + Time × Time. The rationale is that since a coupled system of first-order linear differential equations can fit an exponential function, a log-transformed stock should be predicted well by just a linear term if linearity is a reasonable assumption. We found that, for only 4 out of 11 log-transformed stocks, the Akaike information criterion (AIC) was more than 2 points lower, that is, better, when adding a squared time predictor (less than 2 AIC points difference is not considered meaningful[56]). Furthermore, corresponding scatter plots for the causal links (example provided Figure 8) indicate a relatively low signal-to-noise ratio in the data, suggesting that linear functions are a reasonable approximation and that more, or higher quality, data are likely needed to discern nonlinear effects. Hence, in the absence of clearly discernible nonlinear functional forms in the data, or domain knowledge regarding the causal links' functional forms, describing the causal link with a just a single (linear) term is the most parsimonious option.

We also generated our results with a simulation time of two years (at which point only one log-transformed stock was 2 AIC points lower when adding the squared time predictor, namely morbidity count). This resulted in a similar ranking of simulated intervention effects, with sleep quality, depressive symptoms and hearing loss being the top-ranked risk factors, respectively, and dyslipidemia, blood pressure, diabetes, obesity, and being the lowest-ranked risk factors.

The simulation time was also limited because the ADNI data may contain selective attrition.[57] Even at the 5-year point, only 54% of individuals were still in the study. However, at 2 years 92% of individuals were still in the study and, as stated, the simulated interventions in the SDM with a simulation time of 2 years resulted in a similar ranking of simulated intervention effects that would have resulted in the same conclusions made in this paper. This suggests that the effect of attrition bias is limited within the simulation time of 5 years.

3.4 | Model equations and estimated parameters

3.4.1 | System dynamics equations

Below, we present an example of an algebraic equation (equation 1; sleep quality, measured by the sleep scale from the Medical Outcomes Study [MOS]) and a differential equation (equation 2; depressive symptoms, measured by GDS). The coupled system of equations of all such auxiliaries and stocks constitute the SDM. Stocks and auxiliaries may change over time and therefore have "(t)" next to them to indicate a specific point in time, whereas constants do not. We implemented these equations in Python 3.8 and solved them numerically using the adaptive step size Runge-Kutta method[58] with absolute and relative local error tolerances of 1.4×10^{-8} and 1000 maximum steps for each time-point.

$$\begin{aligned} MOS(t) = & -\theta_{ms,gds} GDS(t) + \theta_{ms,pa} Physical\ activity(t) \\ & + \theta_{ms,cm} Currently\ married + \theta_{ms,age} Age(t) \\ & + \theta_{ms,mg} Gender \end{aligned} \quad (1)$$

$$\begin{aligned} \frac{dGDS(t)}{dt} = & -\theta_{gds,ms} MOS(t) + \theta_{gds,ac} ADAS(t) \\ & + \theta_{gds,cp} Cortisol\ plasma(t) - \theta_{gds,pa} Physical\ activity(t) \\ & - \theta_{gds,cm} Currently\ married + \theta_{gds,age} Age(t) \end{aligned} \quad (2)$$

As seen in these equations, we corrected each auxiliary equation for gender (a constant) and age to adjust for possible confounding effects. Additionally, we adjusted all stock equations for age but not gender after confirming that all stocks, except for BMI, are better predicted in a linear mixed model with an Age × Time interaction but no Gender × Time interaction (as assessed by the AIC).

3.4.2 | Model parameters

In the above equations, $\theta_{i,j}$ represents the parameter that describes the relationship between the feature on the left-hand side of the equation (i) and the feature that it multiplies (j, immediately to its right). Table 2 reports the MAP of each estimated model parameter (averaged over the five training data folds) that corresponds to the strength of the causal link between the origin (incoming) to destination (outgoing) variable. It also reports the 95% highest density interval (HDI), that is, the posterior's narrowest credibility interval that contains 95% of the probability density. In other words: a given parameter lies in this range with 95% probability.

3.5 | Parameter estimation

Due to the substantial missingness in the ADNI data, we estimated auxiliary parameters independently from and prior to estimating the stock parameters. The number of people available to estimate the parameters in the auxiliary equations varied per auxiliary and per data set (see section 3.1) and per data fold (i.e., if an individual was part of the test set fold, it was omitted from the estimation of that fold). After estimating each of the auxiliaries independently, we expressed them in terms of only stocks and constants. That is, we transformed the auxiliaries' equations such that other auxiliaries appearing in these equations were replaced by their right-hand sides until no auxiliaries remained. Consequently, the stocks could be exclusively expressed in terms of other stocks and constants, making the equations used for the estimation of the stocks order-independent. This allowed for the vectorization of the equations, which substantially improved computational efficiency."

3.5.1 | Model without prior

The parameters were independently estimated for each auxiliary equation by ordinary least squares regression on the baseline data (as longitudinal data points were unavailable for many auxiliaries). The auxiliary parameters were then fixed to their mean point estimates in the SDM equations, after which the remaining 88 parameters of the stocks were simultaneously estimated. To estimate the stock parameters, we utilized the available longitudinal data of all individuals in the training set of each data fold. We first used the Levenberg-Marquardt algorithm to obtain the set of parameters corresponding to the lowest sum of squared residuals for each training data fold. After observing parameter identifiability issues and implausible simulated intervention effects, we resorted to Bayesian methods to incorporate prior knowledge regarding the parameter's polarities from the CLD.

3.5.2 | Model with prior

Bayesian parameter estimation is an approach to statistical inference that uses available background knowledge about model parameters in the form of a prior distribution, which is updated using the available data in the form of a likelihood function to obtain a so-called posterior distribution using Bayes' theorem.[59]

We formulated prior distributions for all uncertain parameters by (1) soliciting prior expert knowledge on the link polarities in the CLD and (2) further improving parameter identifiability through the regularizing property of normal priors.[59] For parameters based on CLD links, we selected a half-normal prior centered around 0 with standard deviation $\sigma = 0.5$, which has zero probability for parameter values with a sign other than the CLD's link polarities. For non-CLD parameters, that is, intercepts and parameters involving age and gender, we selected a normal prior with standard deviation $\sigma = 0.5$. This means that approximately 95% of the probability density of the standardized coefficients falls within the interval of [-1, 1] (normal prior) and [0, 1] (half-normal prior). We consider this a valid assumption because the largest absolute parameter in the model without prior was 0.54, and most other parameters were much smaller than that. For the prior of the standard deviation of the normal likelihood, we also utilized a half-normal distribution after checking that the residuals were generally symmetrically centered around the mean and had homogeneous variance over time.

To estimate the model parameters, we first used the No-U-Turn sampler[60] with 1000 samples and 2000 burn-in samples for each auxiliary. Next, we checked the results for convergence using the R^ statistic. Then, we fixed the auxiliary parameters to their mean posterior values and used the Affine Invariant MCMC ensemble sampler to estimate the stock parameters.[61,62] As recommended for practical purposes in this high dimensional problem, we first identified a high probability point using the Trust Region Reflective method, which allows for bounded optimization to find the maximum likelihood point in the parameter space bounded to the sign of the CLD links. As this procedure can get stuck in local minima, we repeated the procedure 10 times with different random initial values. We then initialized the sampler uniformly around the set of parameter values with the lowest sum of squared residuals. Likewise, the standard deviations of the parameters for the likelihood functions were initialized around the root mean squared residual for each stock. Also, as recommended, we used many walkers: 10 times the number of parameters for a total of 880. We collected 300 samples for all walkers, for a total of 4400 posterior samples per data fold after thinning and discarding the first 200 samples per walker. We then checked the acceptance fractions of the walkers, which were 0.2 on average, which is within the recommended range.[62] We also inspected the trace plots to check for convergence of the walkers.

3.6 | Model behavior

3.6.1 | Simulated trajectories

The simulated trajectories of all stocks for a randomly selected individual are shown in Figure 9 for 1000 posterior predictive samples. As can be seen, the posterior trajectories followed the general trend (positive/negative) of the data points for many of the stocks for this individual. However, some simulated trajectories also diverged from

TABLE 2 The magnitude and uncertainty of the SDM parameters

Origin	Polarity	Destination	MAP	95% HDI
Auxiliary parameters				
ADAS-cog-13	–	Cognitive activities score	0.077	[0.0, 0.14]
Currently married	+	Cognitive activities score	0.08	[0.013, 0.143]
Received education	+	Cognitive activities score	0.252	[0.184, 0.324]
Geriatric depression scale	–	Mediterranean diet score	0.066	[0.006, 0.12]
Currently married	+	Mediterranean diet score	0.057	[0.002, 0.11]
Received education	+	Mediterranean diet score	0.179	[0.116, 0.244]
ADAS-cog-13	–	Physical activity score	0.038	[0.0, 0.094]
Geriatric depression scale	–	Physical activity score	0.116	[0.028, 0.201]
Body mass index	–	Physical activity score	0.126	[0.045, 0.214]
Currently married	+	Physical activity score	0.045	[0.0, 0.104]
Received education	+	Physical activity score	0.023	[0.0, 0.063]
Physical activity score	–	MOS-SS	0.036	[0.0, 0.079]
Geriatric depression scale	+	MOS-SS	0.297	[0.234, 0.363]
Currently married	–	MOS-SS	0.036	[0.0, 0.081]
MOS-SS	+	Cortisol plasma	0.042	[0.0, 0.114]
Geriatric depression scale	+	Cortisol plasma	0.027	[0.0, 0.072]
Currently married	–	Cortisol plasma	0.03	[0.0, 0.081]
Received education	–	Cortisol plasma	0.049	[0.0, 0.113]
Mediterranean diet score	–	TNF-alpha plasma	0.082	[0.0, 0.201]
Physical activity score	–	TNF-alpha plasma	0.082	[0.0, 0.192]
Cortisol plasma	+	TNF-alpha plasma	0.036	[0.0, 0.093]
FDG-PET	–	TNF-alpha plasma	0.027	[0.0, 0.072]
Morbidity score	+	TNF-alpha plasma	0.115	[0.001, 0.22]
Body mass index	+	TNF-alpha plasma	0.07	[0.0, 0.171]
Diabetes	+	TNF-alpha plasma	0.035	[0.0, 0.095]
Physical activity score	+	Hippocampal CBF	0.054	[0.0, 0.153]
WMH volume	–	Hippocampal CBF	0.123	[0.0, 0.264]
Mediterranean diet score	+	Superoxide dismutase plasma	0.175	[0.004, 0.333]
Cortisol plasma	–	Superoxide dismutase plasma	0.125	[0.005, 0.233]
FDG-PET	+	Superoxide dismutase plasma	0.064	[0.0, 0.147]
Smoking	–	Superoxide dismutase plasma	0.07	[0.0, 0.155]
Alcohol abuse	–	Superoxide dismutase plasma	0.085	[0.0, 0.183]
Mediterranean diet score	–	TNF-alpha CSF	0.136	[0.0, 0.288]
TNF-alpha plasma	+	TNF-alpha CSF	0.064	[0.0, 0.149]
Superoxide dismutase plasma	–	TNF-alpha CSF	0.093	[0.0, 0.194]
A-beta CSF	–	TNF-alpha CSF	0.043	[0.0, 0.112]
Traumatic brain injury	+	TNF-alpha CSF	0.101	[0.001, 0.209]
Cognitive activities score	–	DMN connectivity a-p ratio	0.093	[0.0, 0.233]
Physical activity score	–	DMN connectivity a-p ratio	0.054	[0.0, 0.151]
MOS-SS	+	DMN connectivity a-p ratio	0.073	[0.0, 0.2]
FDG-PET	–	DMN connectivity a-p ratio	0.073	[0.0, 0.187]

(Continues)

TABLE 2 (Continued)

Origin	Polarity	Destination	MAP	95% HDI
Stock parameters				
MOS-SS	+	ADAS-cog-13	0.202	[0.055, 0.437]
Hippocampal CBF	–	ADAS-cog-13	0.151	[0.067, 0.225]
DMN connectivity a-p ratio	+	ADAS-cog-13	0.133	[0.064, 0.226]
Hippocampal volume	–	ADAS-cog-13	0.289	[0.13, 0.39]
FDG-PET	–	ADAS-cog-13	0.619	[0.431, 0.723]
Geriatric depression scale	+	ADAS-cog-13	0.121	[0.055, 0.194]
Hearing impairment	+	ADAS-cog-13	0.11	[0.052, 0.183]
FDG-PET	+	Hippocampal volume	0.154	[0.118, 0.222]
Mediterranean diet score	+	FDG-PET	0.243	[0.059, 0.542]
Hippocampal CBF	+	FDG-PET	0.12	[0.037, 0.209]
Superoxide dismutase plasma	+	FDG-PET	0.127	[0.065, 0.221]
TNF-alpha CSF	–	FDG-PET	0.172	[0.09, 0.241]
WMH volume	–	FDG-PET	0.186	[0.06, 0.273]
A-beta CSF	+	FDG-PET	0.351	[0.173, 0.578]
Phosphorylated tau CSF	–	FDG-PET	1.059	[0.908, 1.311]
Smoking	–	FDG-PET	0.162	[0.062, 0.24]
Alcohol abuse	–	FDG-PET	0.134	[0.042, 0.207]
Traumatic brain injury	–	FDG-PET	0.133	[0.057, 0.229]
Mediterranean diet score	–	WMH volume	0.163	[0.048, 0.307]
TNF-alpha plasma	+	WMH volume	0.13	[0.042, 0.212]
Hippocampal CBF	–	WMH volume	0.293	[0.114, 0.517]
Superoxide dismutase plasma	–	WMH volume	0.094	[0.032, 0.191]
A-beta CSF	–	WMH volume	0.047	[0.023, 0.123]
Pulse pressure	+	WMH volume	0.057	[0.039, 0.145]
Diabetes	+	WMH volume	0.049	[0.019, 0.122]
Dyslipidemia	+	WMH volume	0.049	[0.018, 0.133]
Smoking	+	WMH volume	0.055	[0.024, 0.124]
Traumatic brain injury	+	WMH volume	0.037	[0.008, 0.114]
ApoE-4 alleles	+	WMH volume	0.049	[0.008, 0.119]
MOS-SS	–	A-beta CSF	0.103	[0.027, 0.208]
TNF-alpha CSF	–	A-beta CSF	0.157	[0.074, 0.228]
WMH volume	–	A-beta CSF	0.132	[0.053, 0.193]
Diabetes	–	A-beta CSF	0.08	[0.029, 0.184]
Traumatic brain injury	–	A-beta CSF	0.083	[0.03, 0.193]
ApoE-4 alleles	–	A-beta CSF	0.125	[0.061, 0.209]
ADAS-cog-13	+	Functional activities questionnaire	0.839	[0.643, 0.992]
FDG-PET	–	Functional activities questionnaire	0.624	[0.457, 0.767]
Morbidity score	+	Functional activities questionnaire	0.115	[0.047, 0.195]
Motor strength impairment	+	Functional activities questionnaire	0.123	[0.054, 0.217]
Mediterranean diet score	–	Morbidity score	0.331	[0.068, 0.712]
Physical activity score	–	Morbidity score	0.29	[0.038, 0.735]
MOS-SS	+	Morbidity score	0.132	[0.063, 0.235]
TNF-alpha plasma	+	Morbidity score	0.137	[0.049, 0.211]
Body mass index	+	Morbidity score	0.161	[0.082, 0.268]

(Continues)

TABLE 2 (Continued)

Origin	Polarity	Destination	MAP	95% HDI
Diabetes	+	Morbidity score	0.223	[0.027, 0.569]
Dyslipidemia	+	Morbidity score	0.231	[0.077, 0.385]
Currently married	–	Morbidity score	0.172	[0.054, 0.388]
Alcohol abuse	+	Morbidity score	0.104	[0.046, 0.227]
Physical activity score	–	Geriatric depression scale	0.137	[0.066, 0.232]
MOS-SS	+	Geriatric depression scale	0.116	[0.064, 0.214]
Cortisol plasma	+	Geriatric depression scale	0.116	[0.074, 0.22]
ADAS-cog-13	+	Geriatric depression scale	0.149	[0.043, 0.284]
Currently married	–	Geriatric depression scale	0.139	[0.066, 0.229]
Mediterranean diet score	–	Body mass index	0.101	[0.031, 0.174]
Physical activity score	–	Body mass index	0.124	[0.044, 0.195]
Mediterranean diet score	–	Pulse pressure	0.158	[0.06, 0.226]
Physical activity score	–	Pulse pressure	0.198	[0.055, 0.356]
Cortisol plasma	+	Pulse pressure	0.169	[0.053, 0.441]
Body mass index	+	Pulse pressure	0.13	[0.063, 0.247]
Currently married	–	Pulse pressure	0.108	[0.038, 0.192]
MOS-SS	+	Phosphorylated tau CSF	0.173	[0.063, 0.281]
Superoxide dismutase plasma	–	Phosphorylated tau CSF	0.187	[0.073, 0.267]
TNF-alpha CSF	+	Phosphorylated tau CSF	0.202	[0.068, 0.24]
A-beta CSF	–	Phosphorylated tau CSF	0.117	[0.06, 0.198]
Traumatic brain injury	+	Phosphorylated tau CSF	0.104	[0.056, 0.193]
ApoE-4 alleles	+	Phosphorylated tau CSF	0.079	[0.012, 0.167]

The polarity is the sign of a given causal link from the causal loop diagram.[4] The parameters are constrained to [0, ∞] by the half-normal prior distribution, which is added or subtracted in the corresponding equation based on the causal link's polarity. For example, the causal link from BMI to pulse pressure has a positive polarity, so its effect is 0.002, whereas the causal link from Mediterranean diet score to pulse pressure has negative polarity, so its effect is -0.002. To adjust for the potentially confounding effects of age and gender, they were also estimated as part of the model (estimates not shown).
Abbreviations: MAP, maximum a posteriori estimate of the parameters; HDI, Highest density interval; DMN connectivity a-p ratio, default mode network connectivity anterior-posterior ratio; CBF, cerebral blood flow; CSF, cerebrospinal fluid.

the data whenever an individual was too dissimilar from the population. For instance, blood pressure, operationalized as pulse pressure (mmHg), decreased only slightly in the simulation (similar to the average population slope, which was only -0.3 per year, Table 1) while the data points for this individual decreased much more (Figure 9). Likewise, FDG-PET and WMH volume increased and decreased in the data of this individual (Figure 9), respectively, while on average these actually decrease and increased over time in the population (Table 1). As shown in Figure 10, when using the MAP parameter values, the stocks' trajectories were similar for all the individuals. As such, we conclude that the SDM predicts similar trends for all individuals and may therefore not fit individual deviations from this trend well.

3.7 | Model validation

3.7.1 | Validation tests

In computational modeling, model validation implies confirming that the model achieves its key objectives. This is commonly done by assessing its structure (structural validation) and behavior (behavioral validation). Hereby it is important to use information that was not already used in the development or estimation of the model.[20,40] In terms of structural validation, the structure of the CLD was critically studied by an interdisciplinary group of neuroscience domain experts and reviewed through a review of scientific literature.[4] Here, we describe two tests for behavioral validation: a behavior pattern test that compared the model simulations to unseen test data and a structure-oriented behavior test that compares the simulated interventions to two sets of validation statements. Given that the SDM's key objective involves simulating interventional what-if scenarios,[20] the structure-oriented behavior test has precedence.

3.7.2 | Behavior pattern test

We assessed the generalizability of the SDM by applying five-fold cross-validation, which provided us with five different training-test set combinations. We estimated three measures of predictive accuracy in

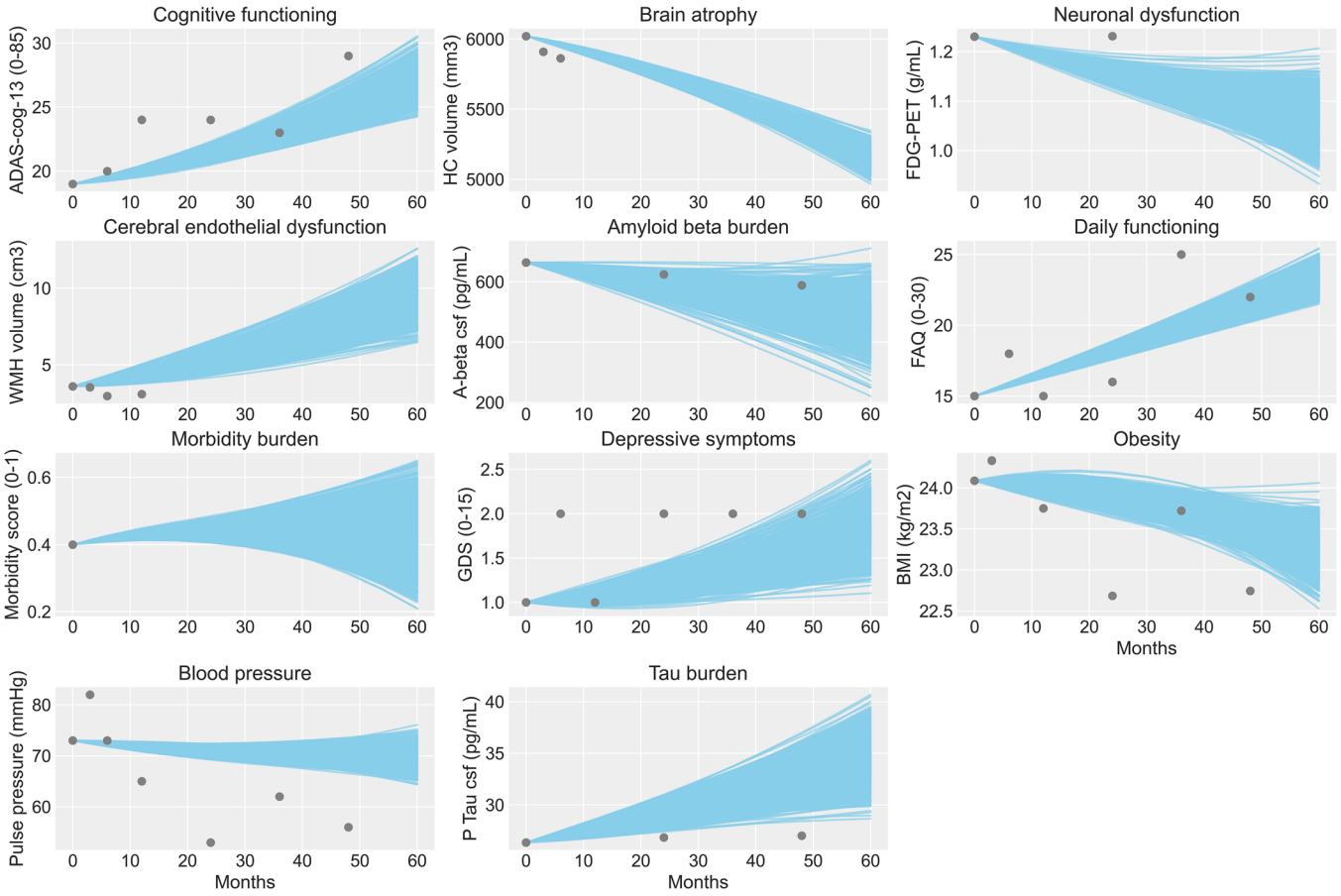

FIGURE 9 Simulated trajectories for all 11 stocks for a randomly selected individual for 1000 posterior predictive samples representing the uncertainty (blue lines). The measurements of this individual are given as data-points (grey dots)

each combination before averaging. Table 3 reports on each of these measures using the MAP estimated parameter values.

The first measure is the mean absolute error (MAE), which measures the average difference between the simulated values and the actual data points in the features' original units and ignores the signs. For instance, the MAE of ADAS-cog-13 was 3.98 points, with a scale of 0-85. This is relatively high considering that, on average, individuals only increased by one ADAS-cog-13 point per year in the data (Table 1). However, the SDM performed 11% better than the reference model and even for models that have point prediction as their principal aim, ADAS-cog-13 trajectories are very difficult to predict in ADNI.[63]

The second measure is the normalized MAE (nMAE), which normalizes the MAE to the mean of the absolute data points. It can therefore be interpreted as a percentage. For instance, the nMAE of hippocampal (HC) volume of 0.03 means that the average deviation between the simulated and actual data points was 3% of the average value of the actual data points, which is very close to the average target value. Conversely, the nMAE of the functional activities questionnaire (FAQ) was 0.71, or 71%, which is very far off from the average target value.

The third measure is a relative error (RE) which compares the RMSE of the SDM to the RMSE of a reference model. Our reference model predicts an increase for each individual in the test set with the average training set slopes for each stock. This RE can also be interpreted as a percentage. For instance, the total RE of the SDM of 0.92 means that the SDM had 8% less error than the reference model. Hence, although the SDM did not generally fit individual deviations from the average well, it did predict 8% better than just predicting the average (as the reference model does) despite only using linear links and fitting 11 stocks simultaneously.

As seen in Table 3, the models with and without prior knowledge had very similar predictive accuracy, having 8% and 9% less error than the reference model, respectively. We also checked the differences between the training and test set errors in each data fold. The training errors were slightly lower than the test errors, but this difference was not statistically significant, suggesting that the SDM did not overfit or underfit the data.

3.7.3 | Structure-oriented behavior test

We conducted a structure-oriented behavior test by comparing observational and interventional meta-analysis evidence to the SDM's simulating interventions rankings. We simulated interventions by alternatively adding or subtracting one standard deviation from the modifiable risk factors. This approach assumes that the population variation

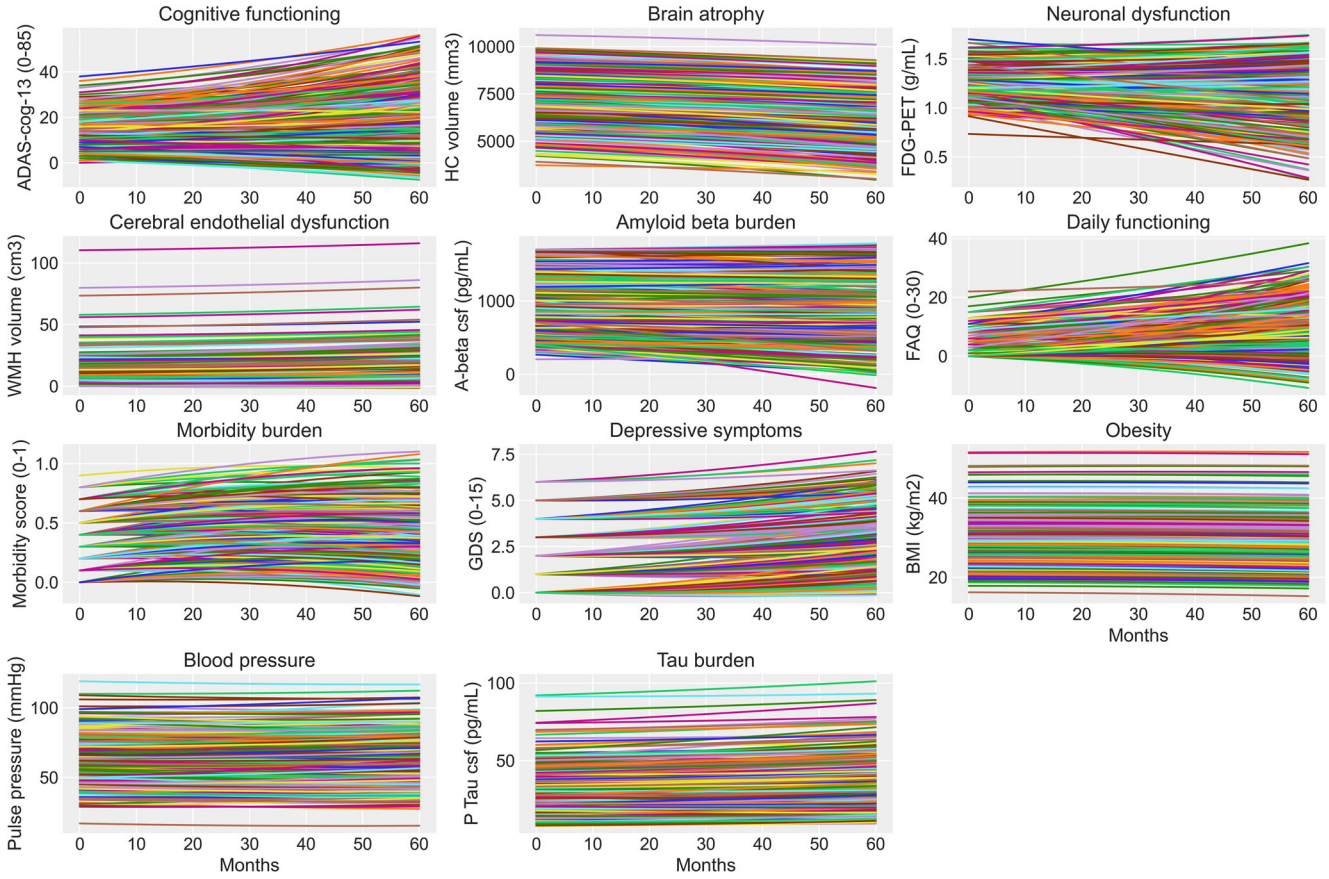

FIGURE 10 Simulated trajectories for all 11 stocks for all individuals for the maximum a posteriori (MAP) parameter values

TABLE 3 Total and per stock predictive accuracy on test data

Type	MAE prior	nMAE prior	RE prior	RE no prior
Total	–	–	0.92	0.91
ADAS-cog-13 (0-85)	3.98	0.30	0.89	0.89
Hippocampal volume (mm^3)	215	0.03	0.76	0.76
FDG-PET (g/ml)	0.05	0.05	1.16	1.15
WMH volume (cm^3)	1.41	0.20	1.0	0.96
A-beta csf (pg/ml)	116	0.11	1.13	0.99
FAQ (0-30)	2.38	0.71	0.8	0.79
Morbidity score (0-1)	0.08	0.22	0.96	0.95
GDS (0-15)	1.13	0.69	1.0	0.99
BMI (kg/m^2)	0.99	0.04	1.0	0.99
Pulse pressure (mmHg)	11.8	0.20	1.01	1.0
P Tau csf (pg/mL)	2.91	0.11	1.06	0.98

Note: Provided for the maximum a posteriori (MAP) parameter values of the SDM with prior. Mean absolute error (MAE) and normalized MAE (nMAE) are given, as well as the relative root mean squared error based on the reference model. For comparison, the latter is also given for the model without prior. No total error is provided for the MAE because the units of the stocks are different and, therefore, not comparable.

represents a certain risk factor's sensitivity to an intervention and that the risk factor effects are independent of other nodes' values. As such, the selected intervention size of one standard deviation was arbitrary, as smaller or larger interventions would have resulted in the same ranking. We considered these assumptions reasonable to obtain a population-level ranking of modifiable risk factors for validation purposes.

The observational validation statements were derived as follows. Livingston et al. report the RR ratios for developing dementia for 12 modifiable risk factors.[12] One of these, air pollution, is not contained

TABLE 4 44 Validation statements based on Livingston et al[12]

Factor	Is greater than factor(s)
Education level	Excessive alcohol use, physical activity, diabetes
Hearing loss	Education level, head trauma, blood pressure, excessive alcohol use, obesity, smoking, social relationships, physical activity, diabetes
Head trauma	Education level, blood pressure, excessive alcohol use, obesity, smoking, social relationships, physical activity, diabetes
Blood pressure	Excessive alcohol use, physical activity, diabetes
Obesity	Excessive alcohol use, physical activity, diabetes
Smoking	Excessive alcohol use, physical activity, diabetes
Depressive symptoms	Education level, head trauma, blood pressure, excessive alcohol use, obesity, smoking, social relationships, physical activity, diabetes
Social relationships	Excessive alcohol use, physical activity, diabetes
Physical activity	Excessive alcohol use
Diabetes	Excessive alcohol use, physical activity

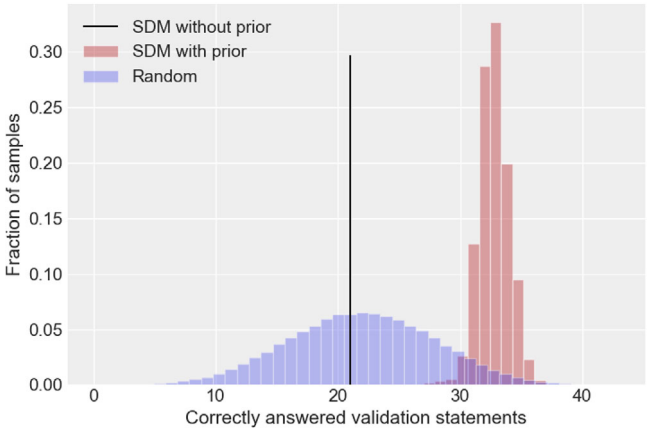

FIGURE 11 Correctly answered observational validation statements compared to 1000 random rankings. The y-axis represents the fraction of rankings that correctly answered a particular number of statements. The SDM with prior was assessed using the intervention effects of 1000 posterior samples, which answered 34 validation statements for the maximum a posteriori (MAP) parameter values, while the model without prior (black line) correctly answered 21 statements

in our model and therefore discarded. Based on the reported RRs, we defined 'is larger than' statements for the remaining 11 risk factors. For instance, depression and hearing loss both have a RR of 1.9, whereas smoking has a RR of 1.6 and excessive alcohol use of 1.2. One of our validation statements thus says that depression should have a larger effect than smoking. Another statement says that depressive symptoms should have a larger effect than excessive alcohol use. Likewise, two other statements say that hearing loss, too, should have a larger effect than smoking and excessive alcohol use. Yet another statement says smoking should have a larger effect than excessive alcohol use. However, no statement concerning depression and hearing loss is defined, as their mean RRs are reported as equal.[12] An overview of all 44 validation statements is given in Table 4. This RR ranking cannot be directly interpreted causally because it is based on a meta-analysis of observational studies. However, we reason that, in the absence of RCT evidence of sufficiently high quality to generate an intervention-based ranking for many of the risk factors, it can nevertheless be used to perform a preliminary assessment of the plausibility of the risk factor ranking generated by the SDM. As such, these observational statements serve as a supplement to the interventional validation statements based on RCTs.

The ranking results are provided in Figure 11. We compared the rankings of the SDM's posterior samples to 1000 random rankings using a Mann-Whitney U test. From the 44 validation statements, the SDM correctly answered between 27 and 37 statements (Mann-Whitney U test: $p < 0.001$), with the MAP correctly answering 34 statements (77%). For comparison, the model that was fitted without prior knowledge regarding the polarity of the causal links answered 21 validation statements correctly (48%). Hence, for the observational validation statements, adding prior knowledge substantially improved the validity of the SDM.

The interventional validation statements were derived as follows. The World Health Organization report on GRADE evidence tables based on systematic reviews for RCT evidence on 12 modifiable risk factors.[12] The evidence was inconclusive for three of these, namely tobacco cessation, hearing loss, and alcohol use disorder. For another four of these, the quality of all the evidence was low (cognitive stimulation, hypertension) or very low (social activity, depression), suggesting that the true effect might be, or even probably is, markedly different from the estimated effect, respectively. Hence, we omitted these factors from our validation statements, leaving five: physical activity, weight reduction, diet, dyslipidemia, and diabetes with moderate quality evidence. For each of these, the standardized mean differences on global cognition (if available, since ADAS-cog-13, the main outcome in this paper, is primarily an index of global cognition) and corresponding average trial durations were reported for specific modalities and, for some factors, specified to cognitive status, that is, mild cognitive impairment (MCI) or cognitively normal (CN). When the overarching conclusion was that no effect could be found, we assumed the estimate to be 0. This was the case for diabetes and dyslipidemia, meaning we could not distinguish between them in the validation statements. For the other factors, we took the most appropriate modality given our selected features, namely aerobic exercise for physical activity, Mediterranean diet for healthy dietary patterns, and lifestyle-based weight reduction for obesity.

The effects were then calculated as follows. For physical activity, the effect of aerobic interventions on cognitive decline was 0.85 (duration: 41.5 weeks) in CN and 0.58 (duration: 24 weeks) in MCI. We thus created a weighted average of these based on the proportion of CN/MCI in the ADNI data (i.e., 56% MCI, Table 1), which resulted in an estimated standardized mean effect of 1.18 per year. No distinction was made between CN and MCI for healthy dietary patterns and

TABLE 5 Nine validation statements based on the World Health Organization[29]

Factor	Is greater than factor(s)
Physical activity	Obesity, healthy dietary patterns, diabetes, dyslipidemia
Obesity	Healthy dietary patterns, diabetes, dyslipidemia
Healthy dietary patterns	Diabetes, dyslipidemia

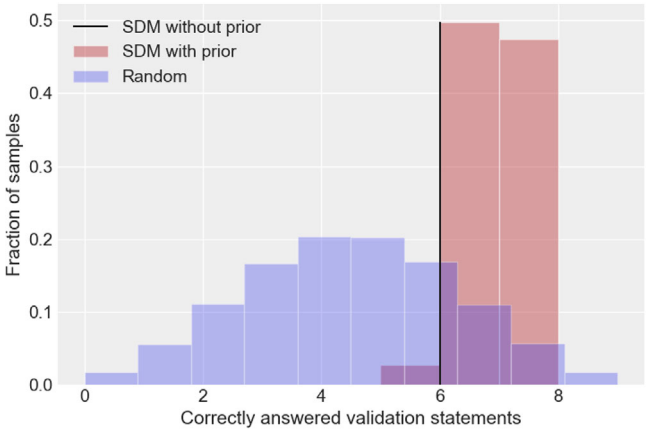

FIGURE 12 Correctly answered interventional validation statements compared to 1000 random rankings. The y-axis represents the fraction of rankings that correctly answered a particular number of statements. The SDM with prior was assessed using the intervention effects of 1000 posterior samples, which correctly answered seven validation statements for the maximum a posteriori (MAP) parameter values, while the model without prior (black line) correctly answered six statements

obesity. Hence, for healthy dietary patterns, the effect of Mediterranean diet was 0.24, normalized by its duration of 4.1 years: 0.05 per year. Finally, for obesity, the effect of weight reduction was not measured on global cognition but rather on attention (effect: 0.44, duration: 20 weeks), executive function (effect: 0), memory (effect: 0.35, duration: 14 weeks), motor speed (effect: 0.16, duration not provided), and language (effect: 0.21, duration: 22 weeks). Hence, assuming an average duration of 0.17 for motor speed, we obtained an effect of 0.65. As can be seen, these estimates are sufficiently different and are thus likely robust against small differences in how they were derived (e.g., averaging the results differently). We then defined nine 'is larger than' statements based on these estimates for the remaining five risk factors. These are provided in Table 5.

The ranking results are provided in Figure 12. From the nine validation statements, the SDM correctly answered between three and eight statements correctly (Mann-Whitney U test: $p < 0.001$), and the MAP correctly answered seven statements (78%). For comparison, the model that was fitted without prior knowledge regarding the polarity of the causal links answered six validation statements correctly (67%). Hence, also for the interventional validation statements, adding prior knowledge improved the validity of the SDM.

ACKNOWLEDGMENTS

This project was partly granted as GEENA-Q-19-595225 by American Alzheimer's Association to M. Olde Rikkert and an unrestricted grant of the research funds from the Devon fund and the Clinical Pharmacologist Professor Dr JM v Rossum-fund, Nijmegen, The Netherlands.

R.Q. and M.O.R. acknowledge the NWO project Social HealthGames (nr. 645.003.002). In addition, R.Q. acknowledges the Netherlands Organization for Health Research and Development (ZonMw) project DINAMICS (nr 531003015), the EU project TO_AITION (nr 848146) and a subsidized project by the Dutch national police (nr 2454972).

Data used in the preparation of this article were obtained from the ADNI and HELIAD studies. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at adni.loni.usc.edu. The HELIAD study was supported by the Grants: IIRG-09-133014 from the Alzheimer's Association, 189 10276/September 8,2011 from the ESPA-EU program Excellence Grant (ARISTEIA) and the ΔΥ2β/οικ0.51657/April 14, 2009 of the Ministry for Health and Social Solidarity (Greece). Data used in the preparation of this article were obtained from the Global Alzheimer's Association Interactive Network (GAAIN), funded additionally by the Alzheimer's Association (Grant no GAAINA-Q-19′595225) at the Mark and Mary Stevens Neuroimaging and Informatics Institute at Keck School of Medicine of USC. The primary goal of GAAIN is to advance research into the causes, prevention, and treatment of Alzheimer's and other neurodegenerative diseases. For up-to-date information, refer to gaain.org. In addition, data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; ElanPharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research Development, LLC.; Johnson Pharmaceutical Research Development LLC.; Lumosity; Lundbeck; Merck Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for NeuroImaging at the University of Southern California.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest. Author disclosures are available in the supporting information.

ORCID

Jeroen F. Uleman https://orcid.org/0000-0001-5050-8505

REFERENCES

1. Fotuhi M, Hachinski V, Whitehouse PJ. Changing perspectives regarding late-life dementia. *Nat Rev Neurol*. 2009; 5(12): 649-658.
2. Hachinski V, Einhäupl K, Ganten D, et al. Preventing dementia by preventing stroke: the Berlin Manifesto. *Alzheimers Dement*. 2019; 15(7): 961-984. doi:10.1016/j.jalz.2019.06.001
3. Sweeney MD, Montagne A, Sagare AP, et al. Vascular dysfunction-The disregarded partner of Alzheimer's disease. *Alzheimers Dement*. 2019; 15(1): 158-167.
4. Uleman JF, Melis RJ, Quax R, et al. Mapping the multicausality of Alzheimer's disease through group model building. *GeroScience*. 2020: 1-15. doi:10.1007/s11357-020-00228-7. Published online.
5. Rollo JL, Banihashemi N, Vafaee F, Crawford JW, Kuncic Z, Holsinger RD. Unraveling the mechanistic complexity of Alzheimer's disease through systems biology. *Alzheimers Dement*. 2016; 12(6): 708-718.
6. Tang Y, Lutz MW, Xing Y. A systems-based model of Alzheimer's disease. *Alzheimers Dement*. 2019; 15(1): 168-171.
7. Fossel M. A unified model of dementias and age-related neurodegeneration. *Alzheimers Dement*. 2020; 16(2): 365-383.
8. Cohen AA, Ferrucci L, Fülöp T, et al. A complex systems approach to aging biology. *Nat Aging*. 2022; 2(7): 580-591. doi:10.1038/s43587-022-00252-6
9. Ahn AC, Tewari M, Poon CS, Phillips RS. The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med*. 2006; 3(6): e208.
10. Butterfield DA, Halliwell B. Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nat Rev Neurosci*. 2019; 20(3): 148-160.
11. Newcombe EA, Camats-Perna J, Silva ML, Valmas N, Huat TJ, Medeiros R. Inflammation: the link between comorbidities, genetics, and Alzheimer's disease. *J Neuroinflammation*. 2018; 15(1): 1-26.
12. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. 2020. Published online.
13. Drzezga A, Becker JA, Van Dijk KRA, et al. Neuronal dysfunction and disconnection of cortical hubs in non-demented subjects with elevated amyloid burden. *Brain*. 2011; 134(6): 1635-1646. doi:10.1093/brain/awr066
14. Johnson GVW, Stoothoff WH. Tau phosphorylation in neuronal cell function and dysfunction. *J Cell Sci*. 2004; 117(24): 5721-5729. doi:10.1242/jcs.01558
15. Mandolesi L, Polverino A, Montuori S, et al. Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. *Front Psychol*. 2018; 9: 509.
16. Yaffe K. Metabolic syndrome and cognitive disorders: is the Sum Greater Than Its Parts? *Alzheimer Dis Assoc Disord*. 2007; 21(2): 167-171. doi:10.1097/WAD.0b013e318065bfd6
17. Kenzie ES, Parks EL, Bigler ED, et al. The dynamics of concussion: mapping pathophysiology, persistence, and recovery with causal-loop diagramming. *Front Neurol*. 2018; 9(4): 203.
18. Hartman YAW, Karssemeijer EGA, van Diepen LAM, Olde Rikkert MGM, Thijssen DHJ. Dementia patients are more sedentary and less physically active than age- and sex-matched cognitively healthy older adults. *Dement Geriatr Cogn Disord*. 2018; 46(1-2): 81-89. doi:10.1159/000491995
19. Gallaway PJ, Miyake H, Buchowski MS, et al. Physical activity: a viable way to reduce the risks of mild cognitive impairment, Alzheimer's disease, and vascular dementia in older adults. *Brain Sci*. 2017; 7(2): 22.
20. Crielaard L, Uleman JF, Chatel BDL, Epskamp S, Sloot PMA, Quax R. Refining the causal loop diagram: a tutorial for maximizing the contribution of domain expertise in computational system dynamics modeling. *Psychol Methods*. 2022. Published online.
21. Farrell S, Stubbings G, Rockwood K, Mitnitski A, Rutenberg A. The potential for complex computational models of aging. *Mech Ageing Dev*. 2021; 193: 111403.
22. Spirtes P, Zhang K. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*. Springer; 2016: 3.
23. Vennix JA. Group model-building: tackling messy problems. *Syst Dyn Rev J Syst Dyn Soc*. 1999; 15(4): 379-401.
24. Crielaard L, Dutta P, Quax R, et al. Social norms and obesity prevalence: from cohort to system dynamics models. *Obes Rev Off J Int Assoc Study Obes*. 2020. Published online.
25. Wakeland W, Kenzie E, A computational model for recovery from traumatic brain injury. In: *Proceedings of the 63rd Annual Meeting of the ISSS-2019* Corvallis, OR, USA. Vol 63;2019.
26. Solomon A, Mangialasche F, Richard E, et al. Advances in the prevention of Alzheimer's disease and dementia. *J Intern Med*. 2014; 275(3): 229-250.
27. Darabi N, Hosseinichimeh N. System dynamics modeling in health and medicine: a systematic literature review. *Syst Dyn Rev*. 2020. Published online.
28. Isaacson RS, Hristov H, Saif N, et al. Individualized clinical management of patients at risk for Alzheimer's dementia. *Alzheimers Dement*. 2019; 15(12): 1588-1602.
29. WHO. Risk reduction of cognitive decline and dementia: WHO guidelines. Published online 2019.
30. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol*. 2018; 14(11): 653-666.
31. Scarmeas N, Anastasiou CA, Yannakoulia M. Nutrition and prevention of cognitive impairment. *Lancet Neurol*. 2018; 17(11): 1006-1015.
32. Mukadam N, Sommerlad A, Huntley J, Livingston G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. *Lancet Glob Health*. 2019; 7(5): e596-e603.
33. Ngandu T, Lehtisalo J, Solomon A, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*. 2015; 385(9984): 2255-2263.
34. Richard E, van Charante EPM, Hoevenaar-Blom MP, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. *Lancet Digit Health*. 2019; 1(8): e424-e434.
35. Vellas B, Carrie I, Gillette-Guyonnet S, et al. MAPT study: a multidomain approach for preventing Alzheimer's disease: design and baseline data. *J Prev Alzheimers Dis*. 2014; 1(1): 13.
36. van Charante EPM, Richard E, Eurelings LS, et al. Effectiveness of a 6-year multidomain vascular care intervention to prevent dementia (preDIVA): a cluster-randomised controlled trial. *The Lancet*. 2016; 388(10046): 797-805.
37. Andrieu S, Guyonnet S, Coley N, et al. Effect of long-term omega 3 polyunsaturated fatty acid supplementation with or without multidomain intervention on cognitive function in elderly adults with memory complaints (MAPT): a randomised, placebo-controlled trial. *Lancet Neurol*. 2017; 16(5): 377-389.
38. Lee WJ, Peng LN, Lin CH, et al. Effects of incorporating multidomain interventions into integrated primary care on quality of life:

a randomised controlled trial. *Lancet Healthy Longev*. 2021; 2(11): e712-e723. doi:10.1016/S2666-7568(21)00248-8
39. Simulating the multicausality of Alzheimer's disease prevention wit… Kumu. Published July 15, 2022. Accessed October 10, 2022. https://jerul.kumu.io/simulating-the-multicausality-of-alzheimers-disease-with-system-dynamics
40. Barlas Y. Formal aspects of model validity and validation in system dynamics. *Syst Dyn Rev J Syst Dyn Soc*. 1996; 12(3): 183-210.
41. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *Lancet Neurol*. 2011; 10(9): 819-828.
42. Bubu OM, Brannick M, Mortimer J, et al. Sleep, cognitive impairment, and Alzheimer's disease: a systematic review and meta-analysis. *Sleep*. 2016; 40(1): zsw032.
43. Cordone S, Scarpelli S, Alfonsi V, De Gennaro L, Gorgoni M. Sleep-based interventions in Alzheimer's disease: promising approaches from prevention to treatment along the disease trajectory. *Pharmaceuticals*. 2021; 14(4): 383.
44. Mukadam N, Anderson R, Knapp M, et al. Effective interventions for potentially modifiable risk factors for late-onset dementia: a costs and cost-effectiveness modelling study. *Lancet Healthy Longev*. 2020; 1(1): e13-e20.
45. Heffernan M, Andrews G, Fiatarone Singh MA, et al. Maintain your brain: protocol of a 3-year randomized controlled trial of a personalized multi-modal digital health intervention to prevent cognitive decline among community dwelling 55 to 77 year olds. *J Alzheimers Dis*. 2019; 70(s1): S221-S237.
46. Zwan MD, Deckers K, Claassen JA, et al. Study design of FINGER-NL: a multidomain lifestyle intervention in Dutch older adults to prevent cognitive decline. *Alzheimers Dement*. 2021; 17: e055136.
47. Sterman JD. All models are wrong: reflections on becoming a systems scientist. *Syst Dyn Rev*. 2002; 18(4): 501-531. doi:10.1002/sdr.261
48. Noble H, Heale R. Triangulation in research, with examples. *Evid Based Nurs*. 2019; 22(3): 67-68.
49. Benveniste H, Liu X, Koundal S, Sanggaard S, Lee H, Wardlaw J. The glymphatic system and waste clearance with brain aging: a review. *Gerontology*. 2019; 65(2): 106-119.
50. GAAIN. The Global Alzheimer's Association Interactive Network. Published online February 2020. http://www.gaain.org
51. Torres-Acosta N, O'Keefe JH, O'Keefe EL, Isaacson R, Small G. Therapeutic potential of TNF-α inhibition for Alzheimer's disease prevention. *J Alzheimers Dis*. 2020; 78(2): 619-626.
52. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review. *J Alzheimers Dis*. 2018; 63(2): 423-444.
53. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol*. 2013; 12(4): 357-367.
54. Bureš V. A Method for simplification of complex group causal loop diagrams based on endogenisation, encapsulation and order-oriented reduction. *Systems*. 2017; 5(3): 46.
55. Hoth KF, Tate DF, Poppas A, et al. Endothelial function and white matter hyperintensities in older adults with cardiovascular disease. *Stroke*. 2007; 38(2): 308-312.
56. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Methods Res*. 2004; 33(2): 261-304.
57. Lo RY, Jagust WJ. Predicting missing biomarker data in a longitudinal study of Alzheimer disease. *Neurology*. 2012; 78(18): 1376-1382.
58. Dormand JR, Prince PJ. A family of embedded Runge-Kutta formulae. *J Comput Appl Math*. 1980; 6(1): 19-26.
59. McElreath R. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC press; 2020.
60. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014; 15(1): 1593-1623.
61. Goodman J, Weare J. Ensemble samplers with affine invariance. *Commun Appl Math Comput Sci*. 2010; 5(1): 65-80.
62. Foreman-Mackey D, Hogg DW, Lang D, Goodman J. emcee: the MCMC hammer. *Publ Astron Soc Pac*. 2013; 125(925): 306.
63. Marinescu RV, Oxtoby NP, Young AL, et al. The alzheimer's disease prediction of longitudinal evolution (TADPOLE) challenge: results after 1 year follow-up. *ArXiv Prepr ArXiv200203419*. 2020. Published online.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Uleman JF, Melis RJF, Ntanasi E, et al. Simulating the multicausality of Alzheimer's disease with system dynamics. *Alzheimer's Dement*. 2023;19:2633–2654. https://doi.org/10.1002/alz.12923