Journal of Biomedical Informatics 145 (2023) 104462

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Exploring the potential impact of multi-factor precision interventions in Alzheimer's disease with system dynamics

Jeroen F. Uleman [a,b,*], René J.F. Melis [b,c], Alfons G. Hoekstra [d], Marcel G.M. Olde Rikkert [a], Rick Quax [b,d], the Australian Imaging, Biomarker and Lifestyle study of Aging and Alzheimer's Disease Neuroimaging Initiative studies[1]

[a] Department of Geriatric Medicine, Radboudumc Alzheimer Center, Donders Institute for Medical Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands
[b] Institute for Advanced Study, University of Amsterdam, Amsterdam, the Netherlands
[c] Department of Geriatric Medicine, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, the Netherlands
[d] Computational Science Lab, Faculty of Science, Informatics Institute, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Precision medicine
Alzheimer's disease
System dynamics
Modifiable risk factors
Multi-factor
Intervention

ABSTRACT

Numerous clinical trials based on a single-cause paradigm have not resulted in efficacious treatments for Alzheimer's disease (AD). Recently, prevention trials that simultaneously intervened on multiple risk factors have shown mixed results, suggesting that careful design is necessary. Moreover, intensive pilot precision medicine (PM) trial results have been promising but may not generalize to a broader population. These observations suggest that a model-based approach to multi-factor precision medicine (PM) is warranted.

We systematically developed a system dynamics model (SDM) of AD for PM using data from two longitudinal studies (N=3660). This method involved a model selection procedure in identifying interaction terms between the SDM components and estimating individualized parameters.

We used the SDM to explore simulated single- and double-factor interventions on 14 modifiable risk factors. We quantified the potential impact of double-factor interventions over single-factor interventions as 1.5 [95% CI: 1.5–2.6] and of SDM-based PM over a one-size-fits-all approach as 3.5 [3.1, 3.8] ADAS-cog-13 points in 12 years.

Although the model remains to be validated, we tentatively conclude that multi-factor PM could come to play an important role in AD prevention.

1. Introduction

Statement of Significance:

Problem	Alzheimer's disease (AD) therapies are not clinically effective, and prevention study results are mixed, suggesting that a multi-factor precision medicine (PM) approach is needed.
What is Already Known	Pilot trials suggest that PM can slow AD but may not generalize to the wider population. Computational models that select the most opportune intervention per person do not yet consider AD's system-wide multicausality.
What This Paper Adds	We applied a systematic method to develop the first system dynamics model for PM in AD. We quantified the potential impact of multi-factor and PM interventions. This initiates a model-based experimental design cycle for iterative model updating.

* Corresponding author at: 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.
E-mail address: jeroen.uleman@radboudumc.nl (J.F. Uleman).

[1] 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. Data was also obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of aging (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO). The AIBL researchers also contributed data but did not participate in the analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

https://doi.org/10.1016/j.jbi.2023.104462
Received 11 February 2023; Received in revised form 9 June 2023; Accepted 26 July 2023
Available online 27 July 2023

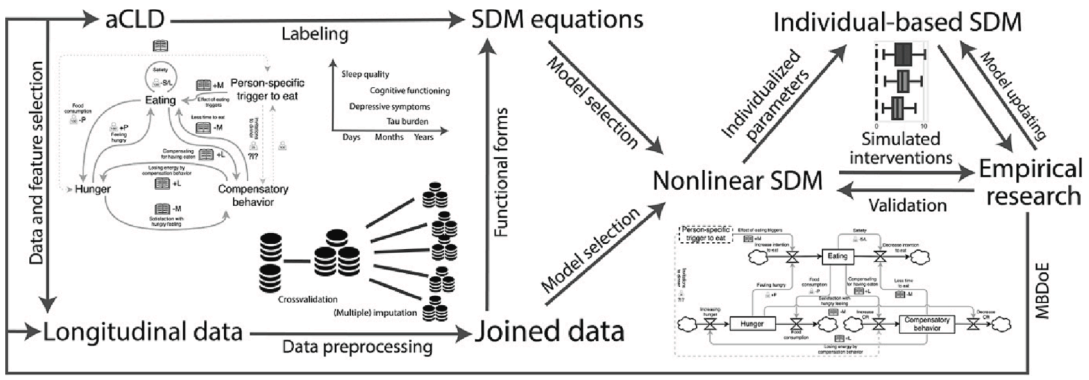

Fig. 1. *Methodology for obtaining a system dynamics model (SDM) for precision medicine (PM), starting from an annotated causal loop diagram (aCLD) and one or multiple longitudinal data sets. Visual example images are also shown. MBDoE = model-based design of experiments. The aCLD and SDM images are examples taken from Crielaard et al. [48] with permission (unrelated to Alzheimer's disease).*

1.1. Precision medicine for Alzheimer's disease

Despite numerous clinical trials based on monocausal paradigms, efficacious treatments that noticeably slow, halt or reverse cognitive decline in Alzheimer's disease (AD) patients are still not realized [1–2]. Among other conclusions, this has led to the increased recognition of the multicausal nature of AD [3–4] and the importance of prevention strategies that target multiple risk factors simultaneously [5]. However, these multi-factor intervention trials have had mixed results [5]. Furthermore, successful compliance to multi-factor interventions depended on intervention complexity and intensity [6]. This suggests that a one-size-fits-all (OSFA) approach may not suffice and that optimal multi-factor interventions differ across persons, e.g., based on age, cognitive status, and dementia risk factors [5,7].

Hence, a personalized or precision medicine (PM) approach may be required that tailors recommendations to the individual [8,9]. In PM, the most opportune intervention is sought for a specific person or subgroup, e.g., through stratification by genetic risk, imaging, or fluid biomarkers [10]. This is thought to lead to less heterogeneity and a better fit of the individual pathophysiology to the mechanisms addressed by the intervention, fewer adverse events, and greater benefit of treatment [11–13].

Nevertheless, it is not yet clear what the added value of PM might be over an OSFA approach to AD. PM pilot studies have shown promising results in preventing cognitive decline, with improvements in cognitive functioning after an intensive (9–18 month) period of personalized prevention recommendations [9,14]. Although this is an encouraging result, it may not fully generalize to a wider population as these pilot studies utilized small samples consisting of highly motivated individuals receiving a large number of recommendations (e.g., 21 per patient on average [9]) with compliance and efficacy that would be difficult to achievable in less motivated individuals [9,14].

To address a wider population, it could be critical to further prioritize preventive interventions on an individual basis. One common approach is to manage AD risk factors to preventively minimize risk. However, although several predictive models for risk assessment have been previously developed for AD [15,16], this approach is based on statistical associations, which may relate poorly to therapeutic effects [17]. Moreover, in other PM approaches for AD, emphasis has been put on genetic risk profiling [17,18], with the non-genetic component of AD remaining underemphasized [19] despite the substantial impact of lifestyle and gene-environment interactions [20,21]. Computational models that integrate the main body of (personalized) multi-factor interventions are thus warranted. As far as we know, such models do not yet exist, although the first promising results have appeared in the literature [22].

1.2. The challenge of developing mechanistic computational models for PM

After sufficient verification and validation [23], and when parametrized with patient-specific data, mechanistic computational disease models can be used to simulate interventional scenarios on an individual basis and support treatment decisions [24,25]. Computational modeling could, therefore, help identify the smallest or least invasive set of interventions predicted to result in the greatest improvements. This set could then be the primary focus of an individual's prevention strategy, improving compliance by intervening on fewer factors.

Although this is a promising vision, developing such validated models is challenging. First, it requires comprehensive background knowledge regarding the multicausal structure of AD's etiology [3], including feedback mechanisms across multiple spatial and temporal scales [22,26–28]. To obtain such background knowledge, numerical data are usually insufficient, meaning that intensive, multidisciplinary teamwork is needed to benefit from other bodies of knowledge, particularly scientific literature and expert knowledge [29]. Second, developing these models requires longitudinal data to quantify this multicausal structure and its dynamics. Finding such data is challenging because the complex and nonlinear relationships between biopsychosocial patient characteristics are rarely sufficiently captured by just one dataset [30]. Moreover, the data that *are* available often have a relatively low signal-to-noise ratio (e.g., substantial measurement error in cognitive functioning questionnaires [31]).

Given the complexity of AD and the sparsity of the data, it may be tempting to follow a reductionist approach and dissect the problem into subproblems to be studied in isolation. Indeed, monocausal paradigms dominate AD research [4], and computational models for AD have focused mainly on single-scale dynamics, such as cellular interactions [32] and biomarker cascades [33–35]. However, although these studies are invaluable for understanding specific AD-related mechanisms, a model for multi-factor PM would arguably require that all causal mechanisms necessary for capturing disease dynamics are included [30], as reductionism is at odds with the required systems approach [27,36].

Hence, there is an urgent need for mechanistic computational models of AD. Although such models may at first lack predictive accuracy at the level of the individual, which would limit their suitability in PM scenarios, their development could initiate a cycle of model-based design of experiments (MBDoE) [37], in which simulations guide empirical research that in turn could be used to update or validate the model. Over several iterations, this might lead to adopting these models for conducting in silico clinical trials [38] and high-fidelity models that can be used in fully personalized PM scenarios.

Table 1
Baseline characteristics of the ADNI and AIBL data features. Miss = missing; CSF = cerebrospinal fluid; CBF = cerebral blood flow; MET = metabolic equivalents of task; IPAQ = International Physical Activity Questionnaire (which may overestimate physical activity compared to accelerometer readings[51]).

		ADNI			AIBL		
aCLD node	**Data feature (units)**	**Mean**	**SD**	**Miss**	**Mean**	**SD**	**Miss**
	Number of participants	1757	–	–	1903	–	–
	Diagnosis (0: Cognitively normal, 1: Mild cognitive impairment)	0.56	–	0%	0.21	0.4	0.1%
	Age (years)	72.9	7.0	0.2%	71.7	7.0	0.1%
	Sex (0: Female, 1: Male)	0.53	–	0%	0.45		0.1%
Auxiliaries							
Diabetes	Fasting glucose (mg/dL)	101	24	30%	94.3	26	38%
Dyslipidemia	Total cholesterol (mg/dL)	194	39	30%	210	43	38%
Experienced stress	Cortisol plasma (ng/mL)	152	54	75%	101	30	54%
Systemic inflammation	TNF-alpha plasma (pg/mL)	8.36	8.0	75%	3.98	2.8	54%
Brain perfusion	Hippocampal CBF (mL/mg/min)	28.5	7.5	92%	–	–	100%
Oxidative stress	Superoxide dismutase plasma (pg/mL)	54.1	36	75%	53.8	24	54%
Neuroinflammation	TNF-alpha CSF (pg/mL)	1.81	1.1	88%	–	–	100%
Neuronal connectivity	Default mode network connectivity anterior-posterior ratio	0.99	0.2	93%	–	–	100%
Physical activity	METs hours per week (IPAQ)	–	–	100%	81.2	75	30%
Healthy dietary patterns	Mediterranean diet score (1–55)	–	–	100%	4.05	1.4	33%
Sleep quality	Pittsburg Sleep Index Questionnaire (0–21)	–	–	100%	4.80	3.3	35%
Stocks							
Cognitive functioning	ADAS-cog-13 (0–85)	13.1	6.9	0.5%	–	–	100%
Neuronal dysfunction	FDG-PET (g/mL)	1.27	0.1	34%	–	–	100%
Amyloid beta burden	Amyloid beta CSF (ng/mL)	1.05	0.5	45%	0.80	0.2	95%
Morbidity burden	Combined morbidity count (0–9)	3.57	1.6	23%	2.01	1.3	61%
Depressive symptoms	Geriatric depression score (0–15)	1.29	1.4	0%	1.22	1.5	17%
Obesity	Body mass index (kg/m2)	27.2	5.0	0.8%	26.4	4.2	27%
Blood pressure	Pulse pressure (mmHg)	59.5	15	0.5%	60.3	15	26%
Tau burden	Phosphorylated tau CSF (pg/mL)	25.5	13	45%	55.6	23	95%
Cerebral endothelial dysfunction	White matter hyperintensities (cm3)	5.66	8.9	35%	8.01	13	86%
Constants							
Social relationships	Currently married (yes/no)	0.75	0.4	0.4%	0.69	0.5	6.5%
Hearing loss	Hearing impairment (yes/no)	0.10	0.3	0%	–	–	100%
Smoking	History of smoking (yes/no)	0.39	0.5	23%	0.41	0.5	16%
Excessive alcohol use	History of alcohol abuse (yes/no)	0.04	0.2	23%	0.12	0.3	56%
Head trauma	Traumatic brain injury (yes/no)	0.05	0.2	23%	0.04	0.2	5.2%
ApoE-4 carriership	ApoE-4 alleles (0–2)	0.48	0.6	1.9%	0.37	0.6	1.8%
Education level	Received education (years)	16.2	2.7	0%	13.0	3.2	25%
Motor function	Motor strength impairment (yes/no)	0.03	0.2	0%	–	–	100%

1.3. Applying system dynamics to precision medicine in Alzheimer's disease

System dynamics is a well-established starting point for combining extensive mechanistic domain knowledge and numerical data into a single computational model. It is closely connected to participatory methods [3,39], which enable co-creation between modelers and domain experts. System dynamics is increasingly applied to medical conditions [40], such as depression [41,42], obesity [43], and concussion [44,45]. Moreover, although early applications of system dynamics to PM may be perceived as reductionist (e.g., focused on only physiological or psychological variables), they demonstrate that the method can have value in PM settings [42,46]. In previous research, we proposed a system dynamics model (SDM) for AD that incorporated 33 risk factors and pathophysiological mechanisms and 148 causal links between them [22]. This SDM was then used to simulate single-factor intervention effects of modifiable risk factors over the studied population, which were validated based on both observational and interventional *meta*-analysis data [20]. However, the model was applied to a population scenario and not tuned for PM. Here, we use it as the foundation for systematically developing the first SDM for PM in AD. Although this SDM cannot yet be validated, we discuss the potential impact of multi-factor PM on AD research based on its preliminary results.

2. Methods

A schematic representation of our method is shown in Fig. 1. The starting point is a causal loop diagram (CLD) which contains additional annotations to facilitate computational modeling, termed an annotated CLD (aCLD) [47]. Subsequently, the aCLD is converted into a nonlinear and, finally, individual-based SDM in a model selection procedure based on longitudinal patient data. The SDM is then used to simulate multi-factor and personalized interventions.

2.1. Annotated causal loop diagram

The SDM is based on an aCLD of sporadic AD (i.e., non-familial AD). The aCLD was developed in a group model building process [39,48,49] involving 15 domain experts [3]. In general, aCLDs contain nodes and directed causal links between them (e.g., X->Y) that can either be positive (more X results in more Y), negative (more X results in less Y), or nonlinear (e.g., a U-shaped relationship), and may form interaction terms [47]. The annotations for nonlinear functional forms and interaction terms will, in our method, be estimated from the data through model selection (Fig. 1). Other annotations include literature references and consensus (or lack thereof) among experts per causal link.

2.2. Data and feature selection

To have data for as many aCLD nodes as possible, we combined data from two studies: the Alzheimer's Disease Neuroimaging Initiative (ADNI), which we also used to develop our previous SDM [22], and the Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging (AIBL). To allow for pooling with ADNI, AIBL utilized tests and protocols that facilitated comparison [50]. Both studies were conducted in a Western context (the United States and Australia) and utilized similar exclusion criteria, such as non-AD dementia, a psychiatric illness, or an uncontrolled medical condition. These criteria are typical for an AD prevention trial population, implying that the SDM can only be applied

to this context. Since the focus of our work is AD prevention, we excluded individuals with diagnosed dementia (or a mini-mental state exam [MMSE] lower than 24 points) at baseline (i.e., the first available time-point). We utilized a simulation time of 12 years as the AIBL data had measurements up to 12 years and ADNI up to 16 years. The baseline characteristics are given for all included features in both data sets in Table 1. More study details are provided in Appendix A.

For the most part, we selected the same data features to operationalize the aCLD nodes as for our previous SDM [22]. Most of these features were present in both data sets, such as the geriatric depression scale (aCLD node: depressive symptoms). However, some features were available in only one of the data sets, such as FDG-PET and ADAS-cog-13 in ADNI and Pittsburg Sleep Index Questionnaire and Mediterranean diet score in AIBL. Finally, we left out three of the nodes from our previous SDM. Since engagement in cognitively demanding tasks (a node with only one outgoing link) was not available in either data set, it was replaced by direct links from its inputs to output. Additionally, brain atrophy, which has only one ingoing and outgoing link, was replaced by a direct link [22], and daily functioning, which has no outgoing links and therefore does not impact other nodes in the model, was omitted to limit the number of model parameters.

2.3. Data preprocessing

Before joining the two data sets, we preprocessed them. First, to assess the (out-of-sample) predictive accuracy of the model, select hyperparameters (as part of the model selection procedure) and estimate the SDM's parameters, we performed holdout cross-validation[i] and split both data sets into training (60% of individuals), holdout (20%), and test (20%) subsets, respectively (see Appendix B for more details).

Second, to account for distributional shifts due to platform and assays differences, we turned phosphorylated tau (CSF), amyloid-beta (CSF), superoxide dismutase (plasma), tumor necrosis factor-alpha (plasma), cortisol (plasma), total cholesterol (plasma) into Z-scores using the baseline means and standard deviations of the training subsets. Although some of these features were skewed, we considered them to overlap sufficiently as Z-scores between the data sets for comparison (corresponding histograms are provided in Appendix C).

Finally, we joined the subsets of the two data samples and applied multiple imputation [52,53] using multivariate imputation by chained equations [54] to impute missing baseline values in the 28 SDM features. Since the SDM requires complete baseline data, this imputation allowed us to run simulations for all individuals despite the considerable missingness (Table 1). To optimize the imputation model [52], we also added eight additional features that were widely available in both data sets: hippocampal volume, MMSE, cognitive status, age, sex, systolic blood pressure, diabetes, and dyslipidemia, which each had Pearson correlations $r>0.22$ ($r=0.45$ on average) to at least one of the SDM features. For example, MMSE, cognitive status, and hippocampal volume had $r=-0.83$, $r=0.72$, and $r=-0.57$ with ADAS-cog-13, respectively. Together with the SDM features, these additional features facilitated the imputation of ADAS-cog-13 in the AIBL data. We used the training set to fit the imputation model. We assess the accuracy of this imputation in Appendix C. We then generated 5 data sets[ii] for the training set to account for the uncertainty in the imputation model [52,56]. The holdout and test sets were imputed only once (with their most likely values) as they were not used to estimate parameter uncertainties. During the multiple imputation, we kept the persons with dementia in the joined data set to maximize the available information. After the multiple imputation, the joined data had a total sample size of N=4637. We then dropped the persons with dementia or MMSE<24 at baseline (N=977), decreasing the sample size to N=3660. Consequently, the test set had a size of N=727 (19.9%), whereas the holdout set had a size of N=723 (20%). We did not impute the longitudinal data because the average change over time in the stocks of people who dropped out of the studies earlier did not differ substantially from people that stayed in the data longer (Appendix C).

2.4. Labeling

Before turning the aCLD into an SDM, a necessary step is labeling the aCLD nodes into either stocks (gradually changing), auxiliaries (fast changing), or constants [47]. Stocks are variables that accumulate over time and change gradually as a function of their inputs (e.g., body mass index), auxiliaries are variables that can change rapidly as a function of their inputs (e.g., inflammatory signaling), and constants keep fixed values over time and only differ between persons (e.g., genotype). In this labeling process, we made the same time-scale assumptions as our previous SDM [22]. The only exceptions are that we operationalized diabetes and dyslipidemia as auxiliaries rather than as constants, measured by fasting glucose (mg/dL) and total cholesterol (mg/dL), respectively. All remaining constants changed in no>10% of the individuals over 12 years. The labeling process is described in more detail in Uleman et al. [22] and Crielaard et al. [47].

2.5. Model selection

Given the labeled aCLD, a system of differential–algebraic equations can be defined in which the differential equations correspond to the stocks and the algebraic equations to the auxiliaries. The ingoing links to these variables, including the constants, will then appear on the right-hand sides of the equations. Given our limited aCLD annotations, however, there was substantial uncertainty regarding the functional forms of the causal links. Together these functional forms (e.g., interaction terms) comprise a vast space of possible SDM formulations. Hence, an efficient model selection procedure was needed to identify the most appropriate SDM (e.g., that maximizes the overall predictive accuracy).

For this reason, we conducted sparse regression [57]. With sparse regression, the parameters in a nested model[iii] were estimated with L1 regularization [59] to force the forms with little predictive power to zero. We first decided which functional forms should be included. Our previous SDM was already able to fit the trends in the stocks on the population level with only linear terms [22]. Here we add additional interaction terms to allow for expected interaction effects in the system, such as the synergistic effect of multimorbidity [60]. We excluded self-loops (i.e., a stock appearing on its own right-hand side) as adding them resulted in highly complex dynamical behavior that was unwarranted given the monotonicity in the trends of most stocks in the data. Moreover, given the many feedback loops in the aCLD [3], the stocks' derivatives already depend on the stocks themselves indirectly.

Just like in our previous work, we also added age (stocks) and sex (stocks and auxiliaries) as additional linear terms to all equations (see Uleman et al [22] for more details). Second, we conducted sparse

[i] The downside of this method is that part of the data is not used for fitting the model. Alternative approaches such as k-fold cross-validation use all the data but require the model to be fit multiple times, which can be computationally prohibitive for large, non-linear models.

[ii] The traditional recommendation of 5 data sets primarily accounts for the stability of point estimates while (many) more imputation sets might be needed for obtaining stable standard errors (e.g., a number equal to the fraction of missing data, which in our case could be up to 48 sets) [52,55]. Nevertheless, due to the computationally intensive and exploratory nature of this work, we stick here to the traditional recommendation.

[iii] This nested model, which contained all possible forms, was estimated in a single procedure, which generally makes sparse regression more efficient than alternative methods for model selection like symbolic regression (e.g., using genetic algorithms [58]). Note that the model selection here thus includes parameter estimation.

regression for each auxiliary equation independently because the algebraic equations depend instantaneously on their input and can thus be treated as cross-sections. However, the coupled differential equations used for stocks containing 184 interactions and 74 linear terms should be estimated simultaneously. To improve the efficiency of the model selection, we first fit an SDM with only linear terms and used it to conduct a preselection of the interaction terms using Sobol sensitivity analysis [61] (see Appendix D), leaving a total of 155 stock parameters to estimate. Finally, we fixed the identified auxiliary parameters to their estimated values in the stocks' equations to proceed with the model selection.

The SDM equations were implemented in Python 3.8 and solved numerically using the adaptive step size Runge-Kutta method [62]. The parameters were estimated for each of the five imputed training sets using a mean squared error cost function with L1 regularization with the Trust Region Reflexive method [63]. To improve the identifiability of the model, the parameters for the linear terms were restricted to the polarity of the corresponding causal links (i.e., $[0, \infty]$ for positive and $[-\infty, 0]$ for negative polarity). All other parameters were allowed to range between $[-\infty, \infty]$. This procedure was repeated several times to select an optimal hyperparameter, λ, to determine the strength of the L1 regularization for each procedure (auxiliaries and stocks estimation) and imputed training set (see Appendix D for more details). We obtained a measure of uncertainty over the parameters with Laplace approximation, which uses the 2nd-order derivative of the cost function for the parameters (i.e., the Hessian) at the optimum to approximate the covariance matrix (assuming a unimodal parameter distribution). Finally, we pooled the five resulting parameter estimates (i.e., one based on each imputed training set) to obtain the average (mean) parameters and covariance matrix using Rubin's rules [53]. For example, for the parameter estimates, this amounted to simply averaging the estimates over the imputed datasets. The resulting parameter values are provided in the supplementary materials table.

2.6. Individualized parameters

At this point, we have estimated a single set of parameters for the whole population. Note, however, that the interaction terms already stratify the population into subgroups (e.g., smokers and non-smokers) [47], meaning that the nonlinear model can already make different predictions for specific individuals. Still, the differences between individuals may be even more granular and not fully captured by the interaction terms.

Therefore, we also estimated individual parameters for each individual in the test set. That is, we repeated the SDM parameter estimation but, rather than fitting the same parameters for all individuals, fitted a unique set of parameters for each person based on their specific data points. Due to the sparsity of the data at the individual level (i.e., fewer longitudinal data points than model parameters), however, this introduced a substantial risk of overfitting. We tried to mitigate this in two ways. First, we limited the number of individualized parameters by selecting the most sensitive parameters through Sobol sensitivity analysis [61] (Appendix E). This resulted in 12 parameters that each contributed>1% to the variance in the cost function. Second, we used L2 regularization[iv] to give the individualized parameters a tendency toward the population parameters.

We then fit these parameters to the data of each individual in the test set based on their first 12 months. We used only 12 months to limit the duration a new individual would have to be followed before such individualized parameters could be estimated. This is critical since most AD lifestyle intervention studies span only a few years [5]. To assess the impact using more data points as input, we also estimated individualized parameters based on the first 48 months. Both sets of parameters are provided in the supplementary materials table under the individualized parameters tab. The 12-month individual-based SDM generated all results in this paper.

2.7. Simulated interventions

Once population- and individual-based parameters were estimated for all N=727 individuals in the test set, the SDM was executable and could thus be used to run simulations. First, this enabled simulating an individual's trajectory based on their first 12 months of data and assessing how well this trajectory matches the individual's actual data points (i.e., the out-of-sample predictive accuracy). Second, this enabled simulating 'what-if' scenarios in which the variable of interest (i.e., cognitive functioning) was compared with and without intervention. The difference between the endpoints of these two trajectories (i.e., after 12 years) was considered the effect of the intervention.

Hence, we next sought to assess the potential impact of multi-factor PM by simulating individualized interventions using the 12-month individual-based SDM. Specifically, we intervened on 14 modifiable risk factors: sleep quality, depressive symptoms, diabetes, obesity, blood pressure, excessive alcohol use, smoking, dyslipidemia, healthy dietary patterns, education level, social relationships, hearing loss, head trauma, and physical activity.

First, we assessed whether double-factor interventions have greater effects than single-factor interventions, given the same intervention strength. To this end, we simulated single-factor interventions by adding or subtracting one standard deviation (SD) from a protective or risk factor, respectively.[v] For double-factor interventions, we added or subtracted ½ SD per factor to equalize the total 'intervention strength' in SD units. Then, we conducted these single- and double-factor simulated interventions in all 727 persons and determined which intervention was best for each individual and which was best on average overall.

Second, we compared a one-size-fits-all (OSFA) approach, defined as conducting the overall best intervention in everyone, to a PM approach, defined as conducting the individually best intervention in each person. We hypothesized that the PM approach would have a larger effect on cognitive functioning than the OSFA approach. We also hypothesized that multi-factor interventions would have larger effects than single-factor interventions. To account for the uncertainty in our simulations, we repeated each of these simulated interventions 100 times. Each time, we changed the model parameters by sampling from a multivariate normal distribution based on the stocks' parameter means and approximate covariance matrix (as estimated during the model selection).

2.8. Validation

A critical step in the computational modeling cycle is model validation. This is done by checking whether the model achieves its key objectives. For PM, the SDM's ultimate purpose is to accurately simulate the effect of interventions at the level of the individual. Unfortunately, these individual-level simulated interventions cannot be validated because we lack intervention data on the specific individuals in our data. Hence, our results can be seen as a first exploration of what could be done with SDMs for PM. We, therefore, focus on our secondary objective, which is to assess whether there is at least significant potential for PM in AD. In other words, in this objective, we are not yet interested in

[iv] As opposed to L1 regularization, L2 regularization does not force parameters to some value (e.g., zero) but rather makes the coefficient tend towards this value. L2 regularization adds the weighted sum of the squared parameter values to the cost function whereas L1 regularization adds the weighted sum of the absolute parameter values.

[v] Specifically, we added or subtracted an SD to the initial values of the stocks and constants while adding it as a constant to the auxiliaries. Note that after such an intervention, stocks and auxiliaries were still allowed to change over time according to the model dynamics.

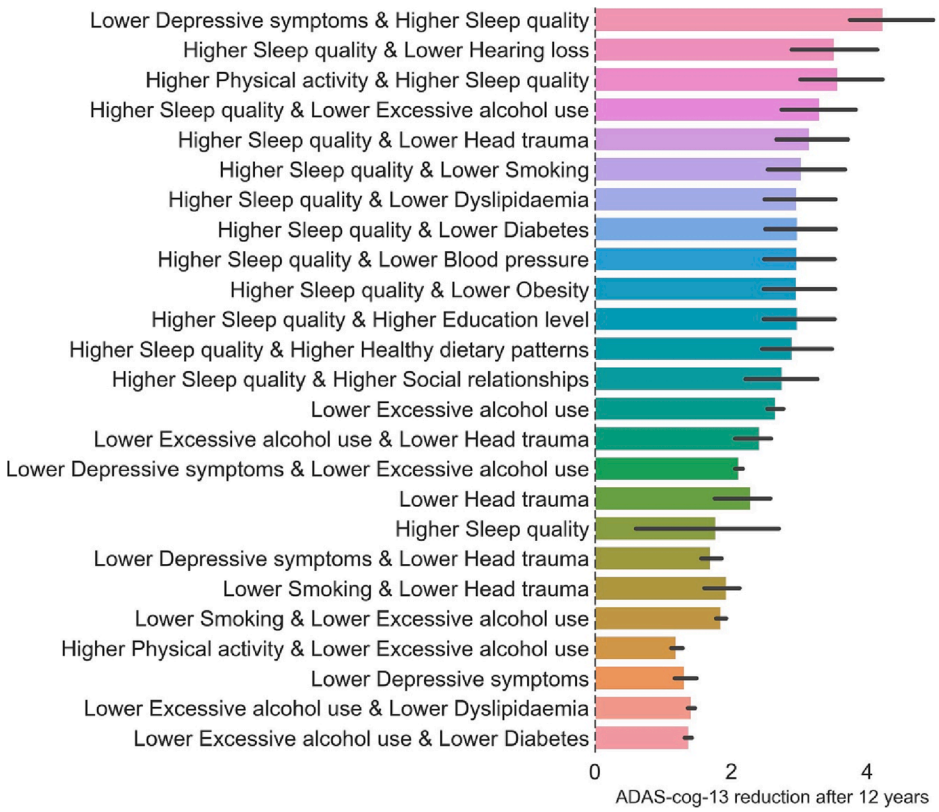

Fig. 2. Top 25 median simulated intervention effects for modifiable factors on cognitive decline (ADAS-cog-13 of all individuals in the test set (N=727). The bars are medians with 95% confidence intervals of the median.

the precise outcome of an intervention for a given individual but in a more general assessment of whether multi-factor and PM interventions tend to differ for different individuals and achieve a significantly larger impact compared to single-factor and OSFA interventions. After all, if this would not be the case, then the PM direction would not seem worthwhile to pursue further, at least for AD.

That said, we did take initial validation steps in our previous work [3,22]. First, after the aCLD was formulated through expert consensus, each link was annotated with at least one supporting reference to scientific literature. Second, in our previous SDM, the simulated interventions were validated against population-level *meta*-analysis results from observational and intervention studies [22]. For more details on the general process of SDM validation, we refer the reader to Crielaard et al. [47].

3. Results

3.1. Out-of-sample predictive accuracy

First, we assessed the predictive accuracy of the SDM by comparing the simulations for each individual in the test set to that individual's actual data points (Appendix F). In sum, we found incremental reductions in the overall test set errors for the different model extensions (linear SDM > nonlinear SDM > 12-month individual-based SDM > 48-month individual-based SDM). This suggests that adding interaction terms and, consequently, the individualized parameters did not overfit the training data. The SDM results thus likely generalize to unseen individuals from a comparable population. Moreover, these improvements in predictive accuracy suggest that using more input data for estimating the individualized parameters was beneficial. Hence, updating an individual's SDM with collected data would result in greater accuracy in predicting that individual's outcomes.

3.2. Multi-factor compared to single-factor interventions

We then assessed the most important risk factors over all individuals in the test set (N=727) by rank-ordering the top 25 single- and double-factor interventions by their median effect per person in Fig. 2. As can be seen, excessive alcohol use, head trauma, sleep quality, and depressive symptoms were the highest-ranked factors for the single-factor interventions. However, double-factor interventions generally ranked higher, with 13 double-factor interventions having greater effects than the top single-factor intervention. This suggests that, given a finite intervention strength, spreading this strength out by simultaneously intervening on multiple risk factors may have a significantly greater effect than concentrating all effort on one of the factors independently.

To quantify the potential impact of this synergistic effect, we compared the intervention on depressive symptoms and sleep quality (top-ranked double-factor intervention, Fig. 2) to the intervention on only excessive alcohol use (top-ranked single-factor intervention, Fig. 2). The median difference between these single- and double-factor interventions based on their median[vi] intervention effects per person was 1.5 [95% CI: 1.0–1.8] ADAS-cog-13 points in favor of the double-factor intervention. In comparison, the median slope in the test data was 0.3 ADAS-cog-13 points per year. This implies that, in this hypothetical scenario, choosing a multi-factor over a single-factor approach would slow cognitive decline by five years on average.

An important observation of this result is that this synergistic effect was highly individualized. For 63% of the individuals, the double-factor intervention had a larger effect, whereas for 37%, it was worse than the single-factor intervention. For the latter 37%, the median benefit of a

[vi] Since the distribution of intervention effects over the individuals was skewed, we used the median as the central tendency for all results with 95% confidence interval of the median based on 10,000 bootstrap samples.

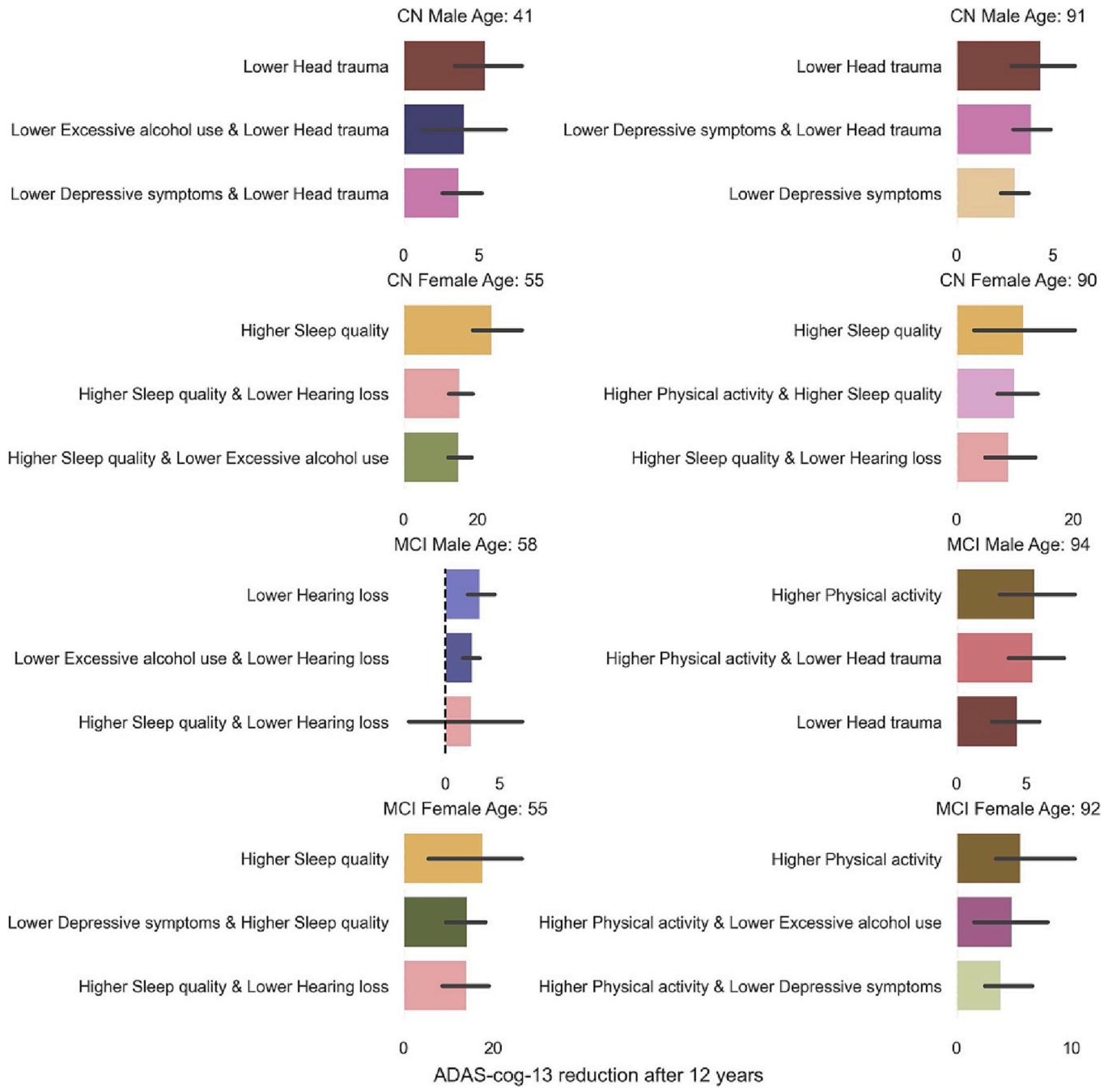

Fig. 3. Top 3 single- and double-factor simulated intervention effects of eight individuals from the test set, namely the youngest and oldest, males and females, without (cognitively normal, CN) and with mild cognitive impairment (MCI) at baseline. The bars are medians with 75% percentile intervals of the median.

single-factor over a double-factor intervention was 3.1 [2.7–3.6 ADAS-cog-13 points in 12 years. Meanwhile, for the former 63%, the benefit of the double-factor over a single-factor intervention was 3.9 [3.4–4.3] ADAS-cog-13 points.

This implies two points that cannot be taken separately. First, performing multi-factor interventions can result in significant improvements compared to single-factor interventions. Second, optimal intervention design depends on the individual. If one would assign the best scoring intervention to each individual (i.e., either a single- or a double-factor intervention), then our model predicts a benefit of 0.6 ADAS-cog-13 points per year on average. We argue that this is a strong motivation for further pursuing PM approaches.

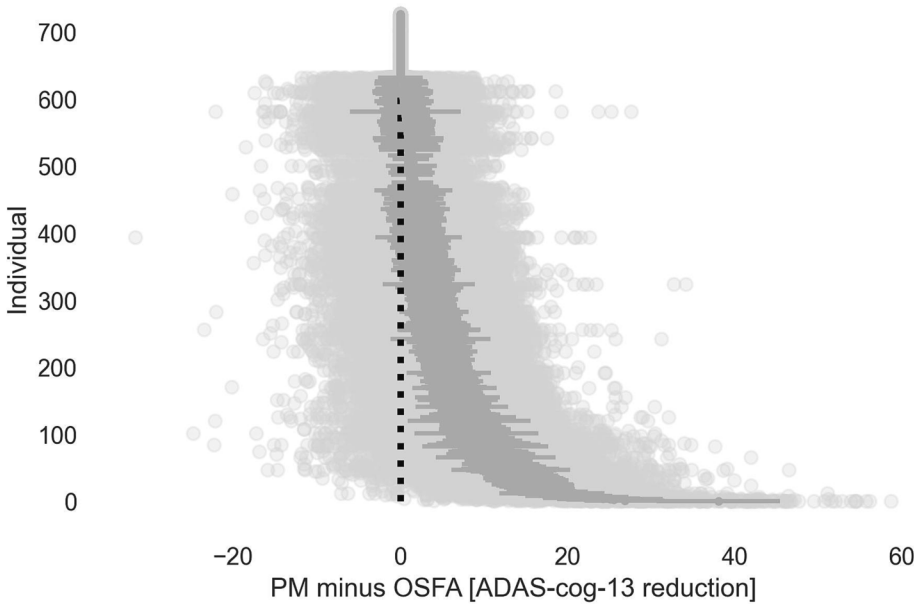

Fig. 4. Difference in ADAS-cog-13 score after 12 years between a one-size-fits-all (OSFA) and a precision medicine (PM) approach to the top-ranked simulated interventions to lessen the cognitive decline in each (N=727) test set individual (y-axis, ordered by their median effect difference) for 100 parameter samples (spread of light grey dots on the x-axis). Each individual's interquartile range of the sample is shown in dark grey. The dotted line represents the null line (i.e., no difference between PM and OSFA). The OSFA and PM interventions were the same for the 13% of individuals in the top.

3.3. Precision medicine compared to one-size-fits-all interventions

To illustrate how the simulated interventions might differ across persons in practice, we plotted the top-ranked simulated interventions for a small selection of individuals. We selected eight individuals from the test set to be diverse over three characteristics, namely: age (youngest and oldest), sex (males and females), and cognitive status (cognitively normal, CN; or mild cognitive impairment, MCI). These characteristics were selected because they are commonly used to stratify patients in clinical studies. Fig. 3 shows the top 3 ranked simulated interventions for these eight individuals. In this figure, the distribution per person reflects the uncertainty, i.e., the 100 parameter samples. As can be seen, the most opportune interventions differ across individuals. For example, the top-ranked factor for the 55-year-old cognitively normal female was sleep quality, whereas, for the 92-year-old female with MCI, the highest-ranked factor was physical activity. Interestingly, only 33% of people had a double-factor intervention ranked highest (see the supplementary materials table). For instance, sleep quality ranked highest in 232 individuals, while sleep quality combined with depressive symptoms ranked highest in only 94 persons. Another observation is that the magnitude of these intervention effects can differ substantially across persons. For instance, these 55- and 92-year-old females had median intervention effects of around 25 and 7 ADAS-cog-13 points, respectively. These differences suggest that the overall preventive potential of interventions on modifiable risk factors may vary wildly between people in clinical practice.

Therefore, we next quantified the potential impact of PM by intervening on the top-ranked factor at the individual level (specific to each individual, Fig. 4) compared to the top-ranked factor at the population level (OSFA; i.e., sleep and depressive symptoms, Fig. 2). The differences between the effects of OSFA and PM are shown in Fig. 4. In 87% of individuals, the PM intervention was better than the OSFA intervention on average. For the remaining 13% of individuals, the PM intervention was equal to the OSFA intervention, meaning there was no difference between their effects. Out of the 87% of individuals who would benefit from PM, the median difference between OSFA and PM based on their median intervention effects per person was 3.5 [3.1, 3.8] ADAS-cog-13 points in favor of a PM approach. This is on top of the median effect of the OSFA intervention (lower excessive alcohol use, Fig. 2), which was 4.5 ADAS-cog-13 points, suggesting that PM might improve the effect of OSFA by almost 78%.

Looking more closely, we can see that for a small fraction of individuals, the difference between PM and OSFA is much larger than the median (Fig. 4). For instance, when considering the 8% of individuals (N=59) with 99% or more of their samples (i.e., light grey dots) falling to the right of the null line in Fig. 4, the benefit of PM over OSFA was 14.6 [12.0, 16.1]. In terms of baseline characteristics, this subgroup differed significantly[vii] from the other test set individuals in having poorer cognitive functioning (3.7 ADAS-cog-13 points, $p=0.0001$), more neuronal dysfunction (0.6 SD, $p<0.0001$), higher amyloid beta burden (0.5 SD, $p<0.0001$), more advanced age (3.2 years, $p<0.001$), and being nearly twice as likely to have MCI (64% compared to 36%). This could be a consequence of there being more risk and cognitive decline to prevent. Indeed, in the data, the subgroup had four times faster cognitive decline rates (1.0 compared to 0.25 ADAS-cog-13 points per year, $p<0.05$). These characteristics could be potentially used to select patients that may benefit most from a PM approach.

4. Discussion

In this paper, we developed an SDM for PM in AD to explore simulated multi-factor and PM interventions and tentatively assess the potential impact of both. In describing this process, we illustrated how SDMs could be systematically developed for PM. As far as we know, the SDM developed here is the first computational model for simulating PM interventions for multi-factor AD prevention, including over a dozen modifiable factors.

An innovative aspect of our method is the data-driven model selection to supplement aCLD annotations. These annotations rely on available background knowledge regarding (nonlinear) functional forms, which are not commonly known by experts. We also developed an individual-based SDM, which improved the overall predictive accuracy. Together, these innovations allowed us to simulate multi-factor PM interventions without background knowledge regarding the functional forms in the system.

Although the present SDM is only the first system dynamics-based effort towards PM in AD and remains to be further validated, we conducted a preliminary explorative study of the potential benefit of multi-factor and PM interventions. We found that intervening on two factors

[vii] assessed with a Kruskal-Wallis test performed on each factor using a significance level with Bonferroni correction of 0.002.

rather than one while spreading the intervention strength over the two factors for comparison resulted in larger intervention effects on average. This suggests that multi-factor interventions might have synergistic benefits. Interestingly, the highest-ranked double-factor intervention involved sleep quality with depressive symptoms rather than excessive alcohol use or head trauma (both ranked higher in Fig. 2). In our previous SDM [22], sleep and depressive symptoms were identified as the potentially most important factors. Our present result suggests that these two factors might be best targeted together.

Although we compared both double-factor and single-factor interventions and PM and OSFA approaches, it is important to consider these as separate aspects that should be assessed independently. On average, multi-factor interventions had greater effects, suggesting that targeting, for instance, depressive symptoms and sleep quality simultaneously in prevention RCTs would be beneficial. However, at the individual level, single-factor interventions frequently ranked the highest, as shown in Fig. 3 and the supplementary materials table. This suggests that while multi-factor interventions may be preferred at the population level, single-factor interventions can sometimes be sufficient in a PM approach.

We thus also compared PM to an OSFA approach, in which PM resulted in a 78% improvement over OSFA on average. Moreover, at the extreme, in the 8% who received the most benefit from PM, the effect was higher than their average yearly cognitive decline rate (14.6 over 12 years is 1.2 compared to 1.0 ADAS-cog-13 point), suggesting that, in these people, cognitive decline could be halted or even reversed. This subgroup differed from the other test set individuals in baseline age, cognitive status and functioning, neuronal dysfunction, and amyloid beta burden. Other characteristics like sex, lifestyle, and ApoE-4 genotype were comparable. This suggests that PM may benefit most those who would decline the most and have a biological predisposition, having relevance irrespective of sex or specific risk profile.

In terms of predictive accuracy, the SDM significantly improved when becoming nonlinear and individual-based, although these improvements were marginal (Appendix F). Indeed achieving high predictive accuracy at the individual level is particularly challenging [30], relies heavily on chance events [64]. and may be especially difficult for cognitive functioning (e.g., ADAS-cog-13 in ADNI) [65]. Nevertheless, we did find that additional improvements in predictive accuracy can be obtained by using more data points to train the model (i.e., 48 instead of 12 months, Appendix F). Unfortunately, using more input data has a practical trade-off since an individual would have to be followed longer before predictions can be made.

Therefore, we propose a cyclic MBDoE approach in which SDMs could propose (precision) interventions and then be iteratively updated based on the collected data. For instance, a nonlinear SDM with population parameters could make a first set of predictions and initiate a series of N-of-1 interventions [66]. In such a study, the individual's features would be measured, and various interventions would be simulated. Based on the ranking of factors specific to the individual, a clinician could prescribe an optimal preventive strategy tailored to their needs. The most critical parameters could then be individualized at subsequent research visits using the collected data. This iterative process allows for the SDM to be updated and would guide the clinician in adjusting the individual's preventive strategies while considering the person's specific abilities and preferences. Conducting trials demonstrating the efficacy of PM approaches compared to standard care [30], as exemplified by the EPINOV trial for personalized epilepsy models, represents a critical future step towards implementing such models in clinical practice.

Limitations of our work include the hypothetical simulated interventions of one SD intervention strength at baseline, which was spread out for double-factor interventions. This assumes that the between-persons SD is informative of how much effort an intervention takes, that this effort is equal across persons, and that the effort can be linearly divided over interventions. These assumptions are hard to validate and may have influenced our results. However, simulated interventions in SD units are also used in other SDM literature [42], and we believe it suffices for providing a first approximation of the potential impact of multi-factor PM. In future works, more realistic interventions could be explored, for instance, by using real intervention units and considering an individual's expected compliance. These interventions could also be extended to more than two factors at a time, e.g., focusing on the highest factors to avoid the combinatorial problem of simulating all possible interventions. Furthermore, while our study emphasizes modifiable risk factors, future work could explore interventions targeting AD pathophysiology more directly, such as the inclusion of amyloid beta and tau immunotherapies.

Another consideration for future work is uncertainty quantification, which is critical for validating PM models [23]. For feasibility reasons, we used Laplace approximation, which assumes a unimodal parameter distribution. We also did not quantify the uncertainty arising from the individualized and auxiliary parameters and utilized an external imputation model with limited imputation sets. Hence, future work might involve a sampling-based multi-level parameter estimation method with Bayesian imputation [67], in which stock and auxiliary parameters are simultaneously estimated with the individualized parameters and missing values. Although such an approach would be optimal, in our experience, it was infeasible given the size of the SDM and may involve several significant challenges (e.g., parameter identifiability and appropriate prior selection). Furthermore, future work could aim to improve the efficiency of the sparse regression procedure. Implementing k-fold cross-validation and multiple imputation in a high-performance computing framework could be part of these efforts. That said, besides computational methods, additional data collection is warranted as well. For instance, sleep quality and cognitive functioning were each measured in only one data set (AIBL and ADNI, respectively), so future longitudinal studies should aim to collect high-quality measures of these features within a single multimodal data set.

A final aspect to consider is the accuracy and comprehensiveness of the causal structure implemented by the SDM. Although various biopsychosocial mechanisms were implemented, not all CLD nodes were included [22], including the glymphatic system's function, which may mediate the effects of sleep quality and head trauma [3,22]. Moreover, while the CLD was developed through the consensus of 15 domain experts and a literature review [3], it might still not be fully accurate. To further enhance the SDM, triangulating the evidence for the CLD with causal discovery methods [68] would be valuable when sufficient empirical data become available. Furthermore, future work could include other relevant features, such as air pollution, that were not yet part of the CLD. Given these limitations and the exploratory nature of this work, we re-emphasize that the present SDM is speculative and that, although we took the first validation steps in our earlier work [3,22], various validation steps remain to be performed.

5. Conclusion

Our exploratory analysis found a considerable potential benefit of multi-factor and PM interventions. The validity of these findings remains to be further tested, for instance, in N-of-1 trials. We developed the SDM using systematic methodological steps to illustrate how SDMs might be developed for PM in other complex disorders. We believe this is an important step towards establishing a systematic, iterative MBDoE process between computational modeling and data collection. It may ultimately lead to reliable system-wide computational models and well-founded personalized and multi-factor in silico interventions whose sum is more than their parts.

CRediT authorship contribution statement

Jeroen F. Uleman: Conceptualization, Methodology, Software, Visualization, Formal analysis, Writing – original draft. **René J.F. Melis:**

Conceptualization, Supervision, Writing – review & editing. **Alfons G. Hoekstra:** Supervision, Writing – review & editing. **Marcel G.M. Olde Rikkert:** Conceptualization, Supervision, Writing – review & editing. **Rick Quax:** Conceptualization, Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding: This project was partly granted as GEENA-Q-19-595225 by the American Alzheimer's Association to M. Olde Rikkert and an unrestricted grant of the research funds from the Devon fund and the Clinical Pharmacologist Prof. Dr. JM v Rossum-fund, Nijmegen, The Netherlands. The authors declare that they have no conflict of interest. Additional funding: RQ and MOR acknowledge the NWO project Social HealthGames (nr. 645.003.002). In addition, RQ 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 AIBL 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. Funding for the AIBL was provided in part by the study partners [Australian Commonwealth Scientific Industrial and research Organization (CSIRO), Edith Cowan University (ECU), Mental Health Research Institute (MHRI), Alzheimer's Australia (AA), National Aging Research Institute (NARI), Austin Health, CogState Ltd., Hollywood Private Hospital, Sir Charles Gardner Hospital]. The study also received support from the National Health and Medical Research Council (NHMRC) and the Dementia Collaborative Research Centres program (DCRC2), as well as ongoing funding from the Science and Industry Endowment Fund (SIEF). The authors acknowledge the financial support of the CRC for Mental Health. The Cooperative Research Center (CRC) program is an Australian Government Initiative. 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.; Elan-Pharmaceuticals, 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 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 Neuro-Imaging at the University of Southern California.

Appendix A. Additional study details

Part of the data we used were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public–private partnership led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The AIBL study group collected the remainder of the data we used. The AIBL study's methodology has been reported previously [69] . The AIBL data is available per request via https://www.aibl.csiro.au.

Appendix B. Holdout cross-validation test

Although sparse regression is efficient compared to alternative methods like genetic algorithms, our procedure remained computationally intensive. Given our reliance on multiple imputed data sets, we selected the holdout method over additional resampling methods like k-fold cross-validation. While utilizing all individuals for training and testing is beneficial, our primary objective was not the estimates' precision. Moreover, the potential bias arising from the holdout method is likely minimal. We assessed this through a comparative analysis between 5-fold cross-validation and the holdout method using the linear SDM. The model parameters and standard errors were estimated using both approaches, and subsequent Z-scores and p-values revealed no significant difference between the estimates (with a significance level of $\alpha=0.05$). Based on these results, we assert that the holdout method was appropriate for our study.

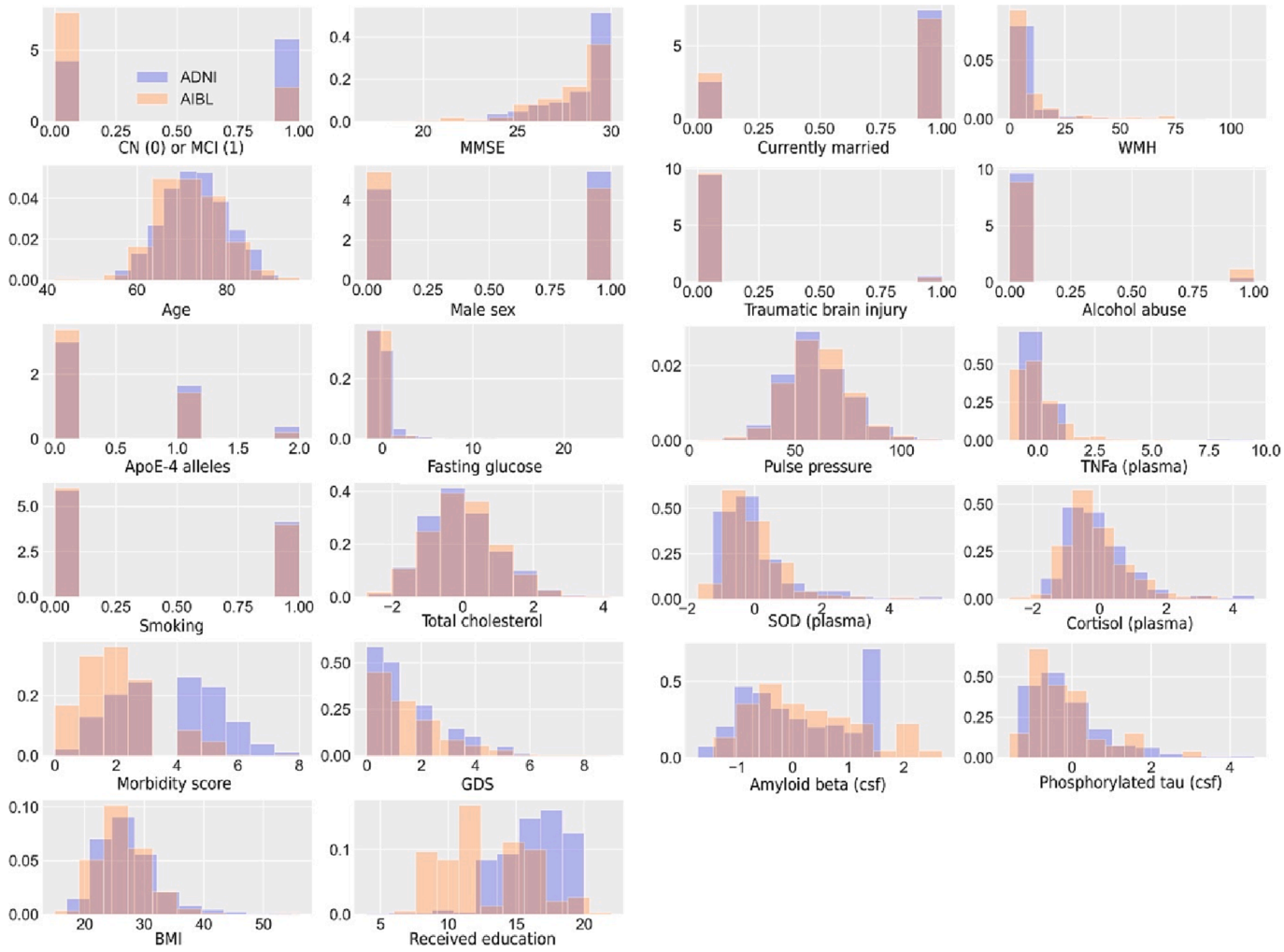

Fig. B.1. Joined training data set before multiple imputation. The dark red color represents the overlap between the distributions.

Appendix C. Imputation details

Feature distributions before imputation

Fig. B.1 shows the distributions of the features used in the imputation model before the data sets (AIBL and ADNI) were joined and imputed. The plotted features were available in both data sets. The features that were available in only one of the data sets (e.g., ADAS-cog-13 and FDG-PET) are not shown. CSF amyloid beta and phosphorylated tau, and plasma cortisol, superoxide dismutase, tumor necrosis factor-alpha, and cholesterol were turned into Z-scores. Since the Z-scores should not contain information from the holdout and test data sets, the two data sets (AIBL and ADNI) were first each split into training-holdout-test sets.

Imputation assessment

To assess the accuracy of the imputation, we conducted a test in which we deleted data points for 25 randomly selected individuals from the training set for each SDM feature. We then averaged the imputed data sets and estimated the root mean squared error (RMSE) between the average and actual imputed values. The RMSE of the standardized data points was lower than one for every feature and 0.74 on average. In comparison, two

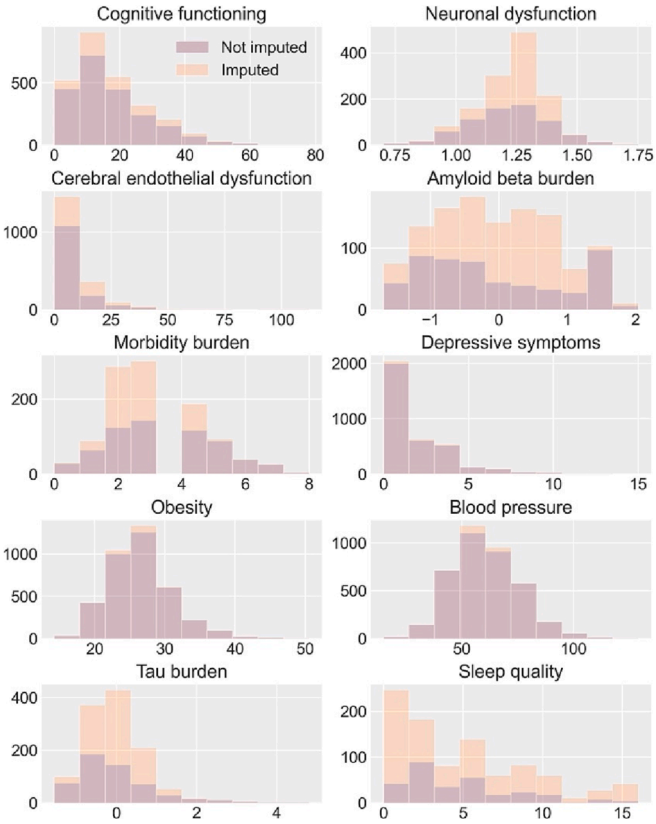

Fig. B.2. Joined training data set before and after imputation for each stock and sleep quality.

samples drawn from a normal distribution with a mean of 0 and an SD of 1 give an RMSE=1.4. Moreover, this error is likely lower for the actual imputation as these additional data points were used to improve the imputation model. Finally, whether the features were available in only one of the data sets or both did not seem to matter. For instance, the RMSEs of cognitive functioning (RMSE=0.68), neuronal dysfunction (RMSE=0.72), and sleep quality (RMSE=0.82) did not differ markedly from features that were available in both data sets (supplementary materials).

To further assess the imputation, Fig. B.2 shows exemplar histograms of the joined training data before and after imputation for sleep quality and the stock nodes. As can be seen, the before- and after-imputation distributions match well.

Longitudinal data test

To test whether people who dropped out of the studies at earlier time points differed from the people who stayed in the studies longer, we stratified

Table C.1
The difference in slopes of the stocks of people who had data points after 36 months (the median, N=1051) compared to people who had data points of 36 months or less (N=1147). A negative slope means that the people who stayed in the data longer had a smaller slope. The slopes are reported in standardized units.

Feature	Slope difference	P-value
ADAS-cog-13	-0.072	0.24
FDG-PET	0.071	0.03
Amyloid beta CSF	-0.006	0.69
Combined morbidity count	0.036	0.21
Geriatric depression score	-0.031	0.98
Body mass index	-0.042	0.37
Pulse pressure	0.025	0.31
Phosphorylated tau CSF	0.027	0.32
White matter hyperintensities	0.001	0.50

the data into individuals who had measurements for ≤36 months (the median, N=1051) and >36 months (N=1147). We found that the average slope (change over time) in the stocks did not differ significantly (Mann-Whitney *U* test, significance level: 0.05) between these two strata in any of the stocks except FDG-PET (Table C.1). Since this procedure involved running multiple tests (increasing the chance of a type 1 error), and the (standardized) slope difference of FDG-PET was only 0.07 standard deviations per year, we concluded that individuals from the strata were sufficiently comparable.

Appendix D. Model selection

Preselection of interaction terms

To improve the efficiency of the model selection procedure, we first fit a linear version of the SDM as a reference (in the same manner as described in the following), in which every causal link was simply a single additive term. We then fixed these linear parameters to their estimated values and added all 184 possible interaction terms to the SDM from which we repeatedly sampled in a Sobol sensitivity analysis [61]. In this sensitivity analysis, we sampled uniformly in the range of [-0.001, 0.001] from all these 184 interaction terms using Saltelli's extension of the Sobol sequence [70]. We took $128 \times (2 + 184) = 23{,}808$ samples and calculated the total Sobol indices for all M model parameters. We then utilized a cut-off for the total Sobol index of 0.0001 to omit interaction term parameters to which the model was insensitive. This resulted in the inclusion of a total of 81 out of the 184 interaction terms. Together with the linear terms, which we re-estimated when we next estimated the interaction terms in the model selection of the nonlinear SDM, this left a total of 155 stock parameters to be selected and estimated. The resulting SDM equations are provided as supplementary materials.

Hyperparameter selection for model selection

Using L1 regularization, the parameters were estimated based on the cost function (mean squared error) and an added penalty term, the sum of absolute parameter values multiplied by a hyperparameter, λ, which controls the strength of the L1 regularization. The effect of λ was assessed using the mean squared error on the holdout data (20% of the individuals). In this way, each parameter was ultimately "selected" – or otherwise set to zero – by the model selection procedure based on its contribution to the predictive accuracy of the SDM in unseen individuals. The λ values for the linear and nonlinear SDM estimation are provided in the supplementary materials table. We repeated the model selection procedure for varying values of λ in increments of 10 until the best value (i.e., the lowest sum of squared errors in the holdout set) was found. A higher λ implies stronger L1 regularization.

Appendix E. Individualized parameters

Sensitivity analysis for individualizing the parameters

As the population parameters were already estimated at this stage, we sampled from the multivariate Gaussian distribution (based on the estimated parameter means and covariance matrix) for the "population" parameters of the nonlinear SDM to incorporate the estimated parameter uncertainty into the procedure. We ran the SDM for $128 \times (2 + \mathrm{M}) = 20{,}096$ samples based on Saltelli's extension for the Sobol sequence for all the M model parameters. We then used a cut-off of 0.01 for the total Sobol indices to select which parameters to individualize. This resulted in 12 parameters that each contributed>1% to the variance in the cost function (i.e., the sum of squared residuals). Due to their relatively large contribution to the cost function, these parameters were individualized.

The most sensitive parameters are shown in Fig. E.1. In general, the effect of age on each of the stocks was more important than the interactions between other factors. This makes sense as advanced age is the most important risk factor of AD and may be understood here as representing many age-related processes that are not included in the aCLD. Future work for the SDM could thus involve MBDoE for elucidating and consequently quantifying some of the latent processes underlying these parameters.

Hyperparameter selection for individualized parameters' estimation

Like L1 regularization, L2 regularization adds a penalty term to the cost function (i.e., mean squared error). The difference is that L2 regularization adds the sum of squared parameter values. Consequently, the parameters are not forced to a specific value but are nevertheless tending to some value (usually zero). The hyperparameter that controls the L2 regularization is typically referred to as α. We selected the α, based on the holdout data, as α=200 for the 12-month individual-based SDM and an α=330 for the 48-month SDM. Starting at zero, α was increased in increments of 10 until a value was reached in which the future data points (i.e., beyond 12 or 48 months and up to 12 years, respectively) were predicted less well. Hence, an α of 200 implies that all lower (0, 10, …, 190) and future values (>210) of α resulted in lower predictive accuracy.

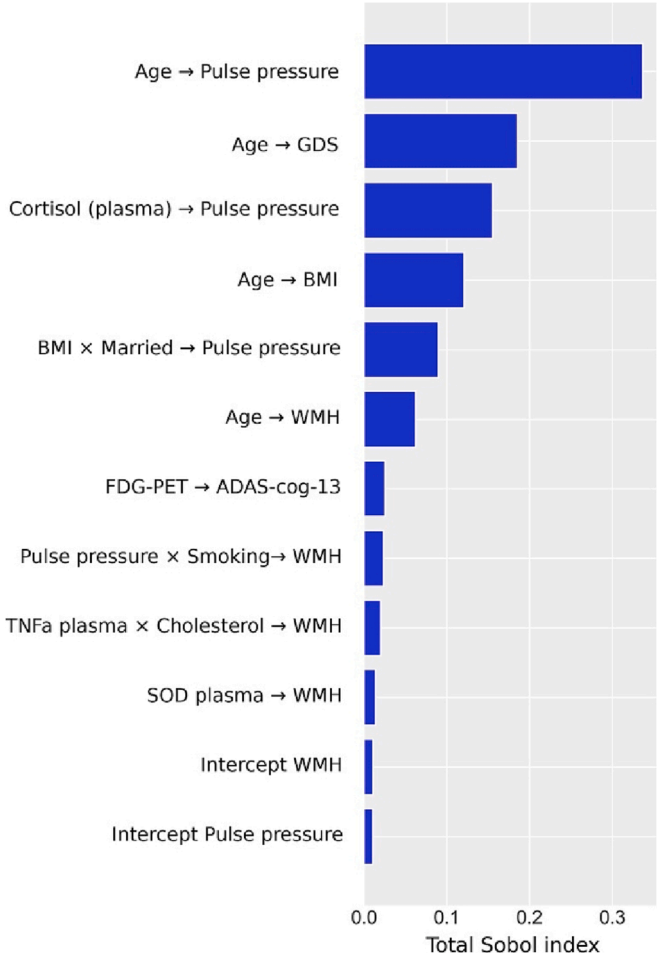

Fig. E.1. Most sensitive twelve parameters based on total Sobol index with a cut-off of 0.01 based on Sobol sensitivity analysis. GDS = Geriatric depression scale, WMH = white matter hyperintensities, SOD = superoxide dismutase.

Appendix F. Predictive accuracy

To assess the out-of-sample predictive accuracy, we estimated the root mean squared error (RMSE) in the individuals from the test set (N=727). The RMSE was calculated for each individual (i) according to equation (F.1), in which M_i is the number of non-missing data points (over all the stocks' time points) per individual, a_{ij} is the individual's actual data point in standardized units and f_{ij} is the corresponding simulated value. A lower RMSE implies greater predictive accuracy.

$$RMSE_i = \sqrt{\frac{1}{M_i}\sum_j^{M_i}\left(a_{ij} - f_{ij}\right)^2} \tag{F.1}$$

The mean and standard deviation over the individual's RMSEs are provided for the different SDM variants in Table F.1. These models were the nonlinear population SDM, the nonlinear individual-based SDMs (with individualized parameters), and a population SDM with only linear terms ('linear SDM') for comparison. The predictive accuracy of the individual-based SDMs could only be assessed with the data points that were not already used to estimate the individualized parameters. Hence, we used the individuals' data points after the first 12 and 48 months to estimate the RMSEs for the 12- and 48-month individual-based SDMs, respectively.

Over all data points, the nonlinear SDM (mean RMSE=0.83) slightly improved compared to the linear SDM (0.84). Adding individualized parameters to the nonlinear SDM further reduced the test set RMSE, namely from 0.88 (linear SDM) and 0.87 (nonlinear SDM) to 0.86 when using the

Table F.1
Test set root mean squared errors (RMSE) of the linear and nonlinear SDMs with population parameters and the individual-based SDMs with individualized parameters. The individualized parameters were fitted to the first 12 or 48 months of data, so for these two individual-based SDMs, the predictive accuracy could only be assessed for the later data points (i.e., >12 months and >48 months, respectively). The RMSE was estimated for every individual over all their available data points (i.e., multiple time points of stocks). The means and standard deviations over individuals are reported.

	Linear SDM	Nonlinear SDM	Individual-based SDM (m12)	Individual-based SDM (m48)
All datapoints	0.84 (0.46)	0.83 (0.46)	–	–
Datapoints >12 months	0.88 (0.48)	0.87 (0.48)	0.86 (0.48)	–
Datapoints >48 months	0.97 (0.51)	0.96 (0.51)	0.95 (0.51)	0.91 (0.50)

first year of data to fit the 12-month individual-based SDM. Moreover, the RMSE reduced further when more input data was used to fit the individual-based SDM, namely from 0.97 (linear SDM), 0.96 (nonlinear SDM), 0.95 (12-month individual-based SDM) to 0.91 (48-month individual-based SDM). To assess the significance of the differences between the model extensions, we conducted two-sided Wilcoxon signed-rank tests over the individual's RMSEs. We compared the linear SDM to the nonlinear SDM and found a difference of 0.004 [95% CI: 0.0004, 0.007] RMSE ($p=0.001$). Next, we compared the nonlinear SDM to the 12-month individual-based SDM and found a difference of 0.008 [0.006, 0.011] ($p<0.001$). Finally, we compared the 12-month to the 48-month individual-based SDM and found a difference of 0.04 [0.02, 0.06] ($p<0.001$).

From these observations, we conclude that adding interaction terms and then individualized parameters significantly, albeit minimally, improved the test error. The nonlinear and individual-based SDMs thus did not result in overfitting on the training data, as expected given the L1 and L2 regularization. Moreover, the relatively larger test error reductions suggest that adding additional data points for estimating the individualized parameters was beneficial, especially for longer-term simulations. Hence, using collected data to update an individual's individualized parameters would result in improved accuracy in predicting that individual's outcomes.

Appendix G. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2023.104462.

References

[1] J. Cummings, G. Lee, P. Nahed, et al., Alzheimer's disease drug development pipeline: 2022, A&D Transl. Res. Clin. Interv. 8 (1) (2022), https://doi.org/10.1002/trc2.12295.

[2] D.S. Knopman, D.T. Jones, M.D. Greicius, Failure to demonstrate efficacy of aducanumab: An analysis of the EMERGE and ENGAGE trials as reported by Biogen, December 2019, Alzheimer's & Dementia. 17 (4) (2021) 696–701, https://doi.org/10.1002/alz.12213.

[3] J.F. Uleman, R.J. Melis, R. Quax, et al. Mapping the multicausality of Alzheimer's disease through group model building, GeroScience. Published online 2020:1-15. doi:10.1007/s11357-020-00228-7.

[4] M. Fotuhi, V. Hachinski, P.J. Whitehouse, Changing perspectives regarding late-life dementia, Nat. Rev. Neurol. 5 (12) (2009) 649–658.

[5] M. Kivipelto, F. Mangialasche, T. Ngandu, Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease, Nat. Rev. Neurol. 14 (11) (2018) 653–666.

[6] N. Coley, T. Ngandu, J. Lehtisalo, et al., Adherence to multidomain interventions for dementia prevention: Data from the FINGER and MAPT trials, Alzheimer's & Dementia. 15 (6) (2019) 729–741, https://doi.org/10.1016/j.jalz.2019.03.005.

[7] G. Forloni, Alzheimer's disease: from basic science to precision medicine approach, BMJ Neurology Open. 2 (2) (2020) e000079.

[8] R.S. Isaacson, C.A. Ganzer, H. Hristov, et al., The clinical practice of risk reduction for Alzheimer's disease: a precision medicine approach, Alzheimer's & Dementia. 14 (12) (2018) 1663–1673.

[9] R.S. Isaacson, H. Hristov, N. Saif, et al., Individualized clinical management of patients at risk for Alzheimer's dementia, Alzheimer's & Dementia. 15 (12) (2019) 1588–1602.

[10] T. Behl, I. Kaur, A. Sehgal, et al., The road to precision medicine: Eliminating the "One Size Fits All" approach in Alzheimer's disease, Biomed. Pharmacother. 153 (2022), 113337, https://doi.org/10.1016/j.biopha.2022.113337.

[11] B. Cholerton, E.B. Larson, J.F. Quinn, et al., Precision Medicine: Clarity for the Complexity of Dementia, Am. J. Pathol. 186 (3) (2016) 500–506, https://doi.org/10.1016/j.ajpath.2015.12.001.

[12] F. Patou, N. Ciccone, J. Thorpe, A. Maier, Designing P4 healthcare interventions for managing cognitive decline and dementia: where are we at? J. Eng. Des. 31 (7) (2020) 379–398, https://doi.org/10.1080/09544828.2020.1763272.

[13] S. Mathur, J. Sutton, Personalized medicine could transform healthcare, Biomedical Reports. 7 (1) (2017) 3–5, https://doi.org/10.3892/br.2017.922.

[14] K. Toups, A. Hathaway, D. Gordon, et al., Precision Medicine Approach to Alzheimer's Disease: Successful Pilot Project, Journal of Alzheimer's Disease. Preprint(Preprint):1–11 (2022), https://doi.org/10.3233/JAD-215707.

[15] Dhana K, Evans DA, Rajan KB, Bennett DA, Morris MC. Healthy lifestyle and the risk of Alzheimer dementia: Findings from 2 longitudinal studies. *Neurology*. 2020; 95(4):e374-e383.

[16] O.J. Schiepers, S. Köhler, K. Deckers, et al., Lifestyle for Brain Health (LIBRA): a new model for dementia prevention, Int. J. Geriatr. Psychiatry 33 (1) (2018) 167–175.

[17] C.L. Berkowitz, L. Mosconi, O. Scheyer, A. Rahman, H. Hristov, R.S. Isaacson, Precision Medicine for Alzheimer's Disease Prevention, Healthcare (Basel). 6 (3) (2018) 82, https://doi.org/10.3390/healthcare6030082.

[18] J.R. Harrison, S. Mistry, N. Muskett, V. Escott-Price, From Polygenic Scores to Precision Medicine in Alzheimer's Disease: A Systematic Review, Journal of Alzheimer's Disease. 74 (4) (2020) 1271–1283, https://doi.org/10.3233/JAD-191233.

[19] Reitz C. Toward precision medicine in Alzheimer's disease. *Annals of Translational Medicine*. 2016;4(6):5-5. doi:10.21037/atm.2016.03.05.

[20] G. Livingston, J. Huntley, A. Sommerlad, et al., Dementia prevention, intervention, and care: 2020 report of the Lancet Commission, Published online, The Lancet, 2020.

[21] A. Eid, I. Mhatre, J.R. Richardson, Gene-environment interactions in Alzheimer's disease: A potential path to precision medicine, Pharmacol. Ther. 199 (2019) 173–187, https://doi.org/10.1016/j.pharmthera.2019.03.005.

[22] J.F. Uleman, R.J. Melis, E. Ntanasi, et al., Simulating the multicausality of Alzheimer's disease with system dynamics, *Alzheimer's & Dementia*. Published online (2023), https://doi.org/10.1002/alz.12923.

[23] M. Viceconti, F. Pappalardo, B. Rodriguez, M. Horner, J. Bischoff, T.F. Musuamba, In silico trials: Verification, validation and uncertainty quantification of predictive models used in the regulatory evaluation of biomedical products, Methods 185 (2021) 120–127, https://doi.org/10.1016/j.ymeth.2020.01.011.

[24] Hampel H, Vergallo A, Perry G, Lista S. The Alzheimer Precision Medicine Initiative. Ghidoni R, Moreira P, eds. *JAD*. 2019;68(1):1-24. doi:10.3233/JAD-181121.

[25] Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA. Causal machine learning for healthcare and precision medicine. *Royal Society Open Science*. 9(8): 220638. doi:10.1098/rsos.220638.

[26] A.A. Cohen, L. Ferrucci, T. Fülöp, et al., A complex systems approach to aging biology, Nat Aging. 2 (7) (2022) 580–591, https://doi.org/10.1038/s43587-022-00252-6.

[27] H.H. Heng, The conflict between complex systems and reductionism, J. Am. Med. Assoc. 300 (13) (2008) 1580–1581.

[28] A.C. Ahn, M. Tewari, C.S. Poon, R.S. Phillips, The limits of reductionism in medicine: could systems biology offer an alternative? PLoS Med. 3 (6) (2006) e208.

[29] Forrester JW. Some basic concepts in system dynamics. *Sloan School of Management, Massachusetts Institute of Technology, Cambridge*. Published online 2009:9.

[30] H. Fröhlich, R. Balling, N. Beerenwinkel, et al., From hype to reality: data science enabling personalized medicine, BMC Med. 16 (1) (2018) 1–15.

[31] J. Feeney, G.M. Savva, C. O'Regan, B. King-Kallimanis, H. Cronin, R.A. Kenny, Measurement Error, Reliability, and Minimum Detectable Change in the Mini-Mental State Examination, Montreal Cognitive Assessment, and Color Trails Test among Community Living Middle-Aged and Older Adults, Journal of Alzheimer's Disease. 53 (3) (2016) 1107–1114, https://doi.org/10.3233/JAD-160248.

[32] W. Hao, A. Friedman, Mathematical model on Alzheimer's disease, BMC Syst. Biol. 10 (1) (2016) 108.

[33] J.R. Petrella, W. Hao, A. Rao, P.M. Doraiswamy, Computational Causal Modeling of the Dynamic Biomarker Cascade in Alzheimer's Disease, Comput. Math. Methods Med. 2019 (2019) 1–8, https://doi.org/10.1155/2019/6216530.

[34] H. Zheng, J.R. Petrella, P.M. Doraiswamy, G. Lin, W. Hao, for the Alzheimer's Disease Neuroimaging Initiative. Data-driven causal model discovery and personalized prediction in Alzheimer's disease, npj Digit Med. 5(1):137 (2022), https://doi.org/10.1038/s41746-022-00632-7.

[35] W. Hao, S. Lenhart, J.R. Petrella, Optimal anti-amyloid-beta therapy for Alzheimer's disease via a personalized mathematical model, PLoS Comput. Biol. 18 (9) (2022) e1010481.

[36] C.G. Langton, Artificial life: the proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, held September,, Los Alamos, New Mexico. 1989 (1987) 6.

[37] G. Franceschini, S. Macchietto, Model-based design of experiments for parameter precision: State of the art, Chem. Eng. Sci. 63 (19) (2008) 4846–4872, https://doi.org/10.1016/j.ces.2007.11.034.

[38] M. Viceconti, A. Henney, E. Morley-Fletcher, In silico clinical trials: how computer simulation will transform the biomedical industry, International Journal of Clinical Trials. 3 (2) (2016) 37–46, https://doi.org/10.18203/2349-3259.ijct20161408.

[39] Vennix JAM. Group Model Building: Facilitating Team Learning Using System Dynamics Field. Wiley; 1996.

[40] Darabi N, Hosseinichimeh N. System Dynamics Modeling in Health and Medicine: A Systematic Literature Review. *System Dynamics Review*. Published online 2020.

[41] N. Hosseinichimeh, A.K. Wittenborn, J. Rick, M.S. Jalali, H. Rahmandad, Modeling and estimating the feedback mechanisms among depression, rumination, and stressors in adolescents, PLoS One 13 (9) (2018) e0204389.

[42] Wittenborn AK, Hosseinichimeh N. Exploring personalized psychotherapy for depression: A system dynamics approach. Branchi I, ed. *PLoS ONE*. 2022;17(10): e0276441. doi:10.1371/journal.pone.0276441.

[43] L. Crielaard, P. Dutta, R. Quax, et al., Social norms and obesity prevalence: From cohort to system dynamics models, *Obes. Rev.*. Published online (2020).

[44] E.S. Kenzie, E.L. Parks, E.D. Bigler, et al., The Dynamics of Concussion: Mapping Pathophysiology, Persistence, and Recovery With Causal-Loop Diagramming, Front. Neurol. 9 (4) (2018) 203.
[45] 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*. ; 2019.
[46] J. Rogers, E.J. Gallaher, D. Dingli, Personalized ESA doses for anemia management in hemodialysis patients with end-stage renal disease, Syst. Dyn. Rev. 34 (1–2) (2018) 121–153.
[47] L. Crielaard, J.F. Uleman, B.D.L. Chatel, S. Epskamp, P.M.A. Sloot, R. Quax, Refining the causal loop diagram: a tutorial for maximizing the contribution of domain expertise in computational system dynamics modeling, Published online, Psychological Methods, 2022.
[48] J.A. Vennix, Group model-building: tackling messy problems, System Dynamics Review: The Journal of the System Dynamics Society. 15 (4) (1999) 379–401.
[49] D.F. Andersen, J.A. Vennix, G.P. Richardson, E.A. Rouwette, Group model building: problem structuring, policy simulation and decision support, J. Oper. Res. Soc. 58 (5) (2007) 691–694.
[50] K.A. Ellis, C.C. Rowe, V.L. Villemagne, et al., Addressing population aging and Alzheimer's disease through the Australian Imaging Biomarkers and Lifestyle study: Collaboration with the Alzheimer's Disease Neuroimaging Initiative, Alzheimer's & Dementia. 6 (3) (2010) 291–296, https://doi.org/10.1016/j.jalz.2010.03.009.
[51] P.H. Lee, Y. Yu, I. McDowell, G.M. Leung, T. Lam, S.M. Stewart, Performance of the international physical activity questionnaire (short form) in subgroups of the Hong Kong chinese population, Int. J. Behav. Nutr. Phys. Act. 8 (2011) 81, https://doi.org/10.1186/1479-5868-8-81.
[52] Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pedersen L, Petersen I. Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*.:10.
[53] D.B. Rubin, Multiple imputation after 18+ years, J. Am. Stat. Assoc. 91 (434) (1996) 473–489.
[54] S. van Buuren, K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, J. Stat. Softw. 45 (2011) 1–67, https://doi.org/10.18637/jss.v045.i03.
[55] P.T. Von Hippel, How many imputations do you need? A two-stage calculation using a quadratic rule, Sociol. Methods Res. 49 (3) (2020) 699–718.
[56] Rubin DB. Multiple imputation for nonresponse in surveys. Published online 1987.
[57] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, PNAS 113 (15) (2016) 3932–3937, https://doi.org/10.1073/pnas.1517384113.
[58] M. Mitchell, An Introduction to Genetic Algorithms, MIT press, 1998.
[59] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, J. Roy. Stat. Soc.: Ser. B (Methodol.) 58 (1) (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[60] A. Calderón-Larrañaga, D.L. Vetrano, L. Ferrucci, et al., Multimorbidity and functional impairment-bidirectional interplay, synergistic effects and common pathways, J. Intern. Med. 285 (3) (2019) 255–271, https://doi.org/10.1111/joim.12843.
[61] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, Comput. Phys. Commun. 181 (2) (2010) 259–270.
[62] J.R. Dormand, P.J. Prince, A family of embedded Runge-Kutta formulae, J. Comput. Appl. Math. 6 (1) (1980) 19–26.
[63] M.A. Branch, T.F. Coleman, Y. Li, A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems, SIAM J. Sci. Comput. 21 (1) (1999) 1–23.
[64] G.D. Smith, Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice, Int. J. Epidemiol. 40 (3) (2011) 537–562.
[65] 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 preprint arXiv:200203419*. Published online 2020.
[66] J.E. Galvin, Advancing personalized treatment of Alzheimer's disease: a call for the N-of-1 trial design, Future Neurol. 13 (3) (2018) 151–160, https://doi.org/10.2217/fnl-2018-0004.
[67] R. McElreath, Statistical Rethinking: A Bayesian Course with Examples in R and Stan, CRC Press, 2020.
[68] P. Spirtes, K. Zhang, Causal discovery and inference: concepts and recent methodological advances, Appl Inform. 3 (1) (2016) 3, https://doi.org/10.1186/s40535-016-0018-x.
[69] K.A. Ellis, A.I. Bush, D. Darby, et al., The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease, Int. Psychogeriatr. 21 (4) (2009) 672–687, https://doi.org/10.1017/S1041610209009405.
[70] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Commun. 145 (2) (2002) 280–297, https://doi.org/10.1016/S0010-4655(02)00280-1.