BRIDGING LONGITUDINAL HEALTH RECORDS AND 3D HUMAN GENERATORS WITH BIOMARKER-GUIDED SIGNED CHOQUET FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Longitudinal health records contain compact, structured measurements, while pretrained 3D human generators operate on high-dimensional visual conditions. We study how to connect these representations when the health cohort has no participant-specific 3D scans. We introduce the signed Choquet twin (SCT), a health digital twin that forecasts body and blood measurements from diet records and turns those forecasts into 3D bodies. SCT predicts each person's expected change from their age, sex and baseline measurements, then adds a diet term built from the 11 components of the Alternative Healthy Eating Index. This term uses a symmetric Choquet integral that assigns a weight to each component and to each pair of components, so every forecast breaks down exactly into the contributions of individual foods and their interactions. A key result is that this diet term is linear in its weights, and the valid weights form two convex cones, which makes estimation clean and exact. Cross-fitted networks first remove the influence of each person's background, and small convex quadratic programs then give the exact weights, with cross-validated shrinkage toward the total diet score. To generate bodies without real scans, SCT learns from generated reference bodies how changes in measurements should shift the input to a frozen image-to-3D generator. On 7,876 participants from two long-running studies, SCT forecasts ten measurements more accurately than regression, gradient boosting and a Choquet baseline. These accuracy gains come from the trajectory networks, while the diet term contributes interpretability. Better diet quality is consistently linked to lower body weight and a smaller waist across all training runs. The total diet score captures most of this effect, and a synthetic cohort with known effects confirms that the model recovers the overall diet effect reliably. The 3D trajectories remain qualitative because measured longitudinal 3D scans are unavailable, so we present SCT as a measurable pipeline from longitudinal health records to explainable diet effects and 3D visualizations, together with a practical route for generating temporal body updates without paired 3D supervision.

1 INTRODUCTION

A health digital twin is a personal model that can be run forward under different behaviours (Björnsson et al., 2020; Katsoulakis et al., 2024); for diet, it forecasts how body and blood measurements change under different eating patterns. A visualized future body can convey risk more directly than numbers, but only if the forecast is trustworthy and shows which diet components drive it. Building such a twin from cohort data requires capturing interactions between overlapping diet components while staying interpretable, letting better and worse diets act in opposite directions, separating the diet effect from the diet-free trajectory driven by age, sex, and baseline measurements, and generating 3D bodies without scans.

We address this with the *signed Choquet twin* (SCT), which models each outcome as its baseline plus a diet-free trajectory plus a diet term. The diet term is a symmetric Choquet integral (Choquet, 1954; Šipoš, 1979) with a 2-additive capacity (Grabisch, 1997) over centred diet components. We show it is linear in 66 weights, with valid weights forming a convex cone and its mirror image, so estimation reduces to partially linear regression (Robinson, 1988; Chernozhukov et al., 2018):

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cross-fitted networks residualize the outcome and diet features, and two small convex quadratic programs give exact weights, shrunk toward the total diet score (Chiuve et al., 2012). For 3D, a linear map learned from 320 generated reference bodies shifts a person's encoding in the frozen generator TRELLIS.2 (Xiang et al., 2025a) by the predicted measurement change.

Our contributions are:

- A diet forecaster built on a symmetric Choquet integral whose capacity is estimated exactly: the diet term is linear in its weights, valid weights form two polyhedral cones, and orthogonalized estimation reduces the fit to convex quadratic programs with a cross-validated pull toward the total diet score. Forecasts split exactly into component and pair parts (Section 3).
- A way to drive a frozen 3D generator from predicted measurements without scans, keeping appearance fixed and changing body volume with predicted weight (Section 3.4).
- An evaluation on 7,876 participants: baselines, ablations of every part of the fit, effects with intervals and tests, measured 3D edits, and a synthetic cohort with known effects (Sections 4–5).

SCT is the most accurate of all compared models (skill 0.249 against 0.227), but the gain comes from the trajectory networks: without diet, or with the total diet score alone, the model is as accurate. The diet term instead gives a stable estimate: a better diet predicts lower weight, waist, blood pressure, and insulin in all 25 runs, with intervals.

2 Related Work

Choquet integrals and capacity learning. Choquet integrals generalize weighted means by assigning a weight to every subset of inputs (Choquet, 1954; Sugeno, 1974; Beliakov et al., 2007), with Shapley and interaction indices that describe the role of each input and pair (Shapley, 1953; Grabisch, 1997; Grabisch and Roubens, 1999). They are widely used in multi-criteria decision aid (Grabisch, 1996; Grabisch and Labreuche, 2010), and the symmetric (Šipoš) integral extends them to inputs of both signs (Šipoš, 1979; Grabisch and Labreuche, 2005). A general capacity must satisfy $n2^{n-1}$ monotonicity constraints (Tehrani et al., 2012), k-additive capacities limit the parameters to subsets of at most k inputs (Grabisch, 1997), and least-squares capacity identification is a quadratic program (Grabisch et al., 2008). We add a signed effect size and embed the 2-additive fit in a semiparametric regression whose other parts are neural networks.

Interpretable and orthogonalized estimation from health records. Additive models (Hastie and Tibshirani, 1986) and neural additive models (Agarwal et al., 2021) give per-feature contributions by design, GA^{2}Ms add learned pairwise terms (Lou et al., 2013), and Shapley-based attributions (Lundberg and Lee, 2017) explain black-box models after training. Our diet term is also additive in components and pairs, but each pair enters through one weight on the fixed form $\min(x_i, x_j)$, so the aggregate is monotone and bounded. Partially linear regression (Robinson, 1988) and double machine learning (Chernozhukov et al., 2018; Nie and Wager, 2021) learn functions of context with cross-fitting and fit the parameter of interest to residuals; here the parameter is a capacity. Diet indices such as AHEI-2010 summarize diet quality with fixed additive rules (Chiuve et al., 2012), and cohort studies relate diet changes to weight change (Mozaffarian et al., 2011).

3D humans from images and measurements. Parametric body models such as SMPL (Loper et al., 2015) and anthropometric predictors trained on scans (McCarthy et al., 2024) map measurements to shape when paired scans exist. Feed-forward image-to-3D generators reconstruct objects and people from a single image (Hong et al., 2024; Tochilkin et al., 2024); TRELLIS (Xiang et al., 2025b) and TRELLIS.2 (Xiang et al., 2025a) generate structured 3D latents conditioned on image features, and text-to-image models (Wu et al., 2025) can render a person with given attributes. Without scans, we learn how measurements move a frozen generator's conditioning from generated reference bodies.

3 Method

Figure 1 shows the forecaster (Sections 3.1–3.3) and the 3D path (Section 3.4).

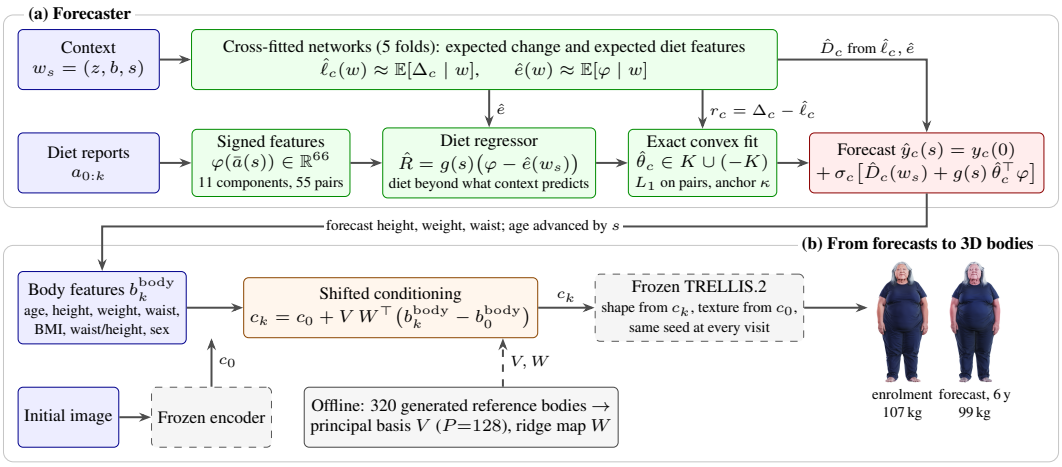

Figure 1: Overview of SCT. (a) Networks predict each outcome's change and the diet features from context; the diet weights $\hat{\theta}_c$ (effect size times 2-additive capacity) exactly solve a convex fit to what context leaves unexplained. (b) Forecast body features shift the encoding of an initial image in a frozen generator (bodies from Figure 3). Dashed: frozen models.

3.1 Problem setup

For each participant, visits occur at elapsed years $0 = s_0 < s_1 < \cdots < s_K$. At visit k we observe the diet $a_k \in [0,1]^d$, the $d = 11$ AHEI-2010 component scores divided by 10, and outcomes $y_k \in \mathbb{R}^C$ with a missingness mask. The $C = 10$ outcomes are height, weight, waist, systolic and diastolic blood pressure, fasting glucose, insulin, high- and low-density lipoprotein cholesterol (HDL, LDL), and the homeostatic model assessment of insulin resistance (HOMA-IR; Matthews et al., 1985). A context vector $z \in \mathbb{R}^{14}$ holds baseline age, sex, study, race and ethnicity indicators, and standardized baseline height, weight, waist, and body-mass index (BMI), and $b \in \mathbb{R}^{20}$ holds the standardized baseline outcomes with missing values set to zero and flagged. We write $w_s = (z, b, s)$ for the context at elapsed time s. The task is to predict y_k for $k \geq 1$ from y_0, z, b, and the diet reports. A report can be read as the diet from that visit until the next (*start* timing, using reports made before visit k) or, since food questionnaires ask about the past year, as the diet before the visit (*report* timing, also using the report at visit k); cross-validation chooses. Forecasts are thus conditional on a diet history, observed or assumed, up to the target visit.

3.2 Forecaster

Outcome equation. For outcome c at elapsed time s, the scaled change $\Delta_c(s) = (y_c(s) - y_c(0))/\sigma_c$ is modelled as

$$\Delta_c(s) = D_c(w_s) + \beta_c\, S_c\big(x(s)\big) + \varepsilon_c, \tag{1}$$

where σ_c is the standard deviation of observed change in the training split, D_c is a diet-free trajectory with $D_c(w_0) = 0$, $\beta_c \in \mathbb{R}$ is the diet effect size, S_c is the diet aggregate defined next, and $\mathbb{E}[\varepsilon_c \mid w_s, x(s)] = 0$.

Signed exposure. Each component is standardized against the training cohort, $\tilde{a}_j = (a_j - \eta_j)/\omega_j$ with η_j and ω_j its mean and standard deviation, so positive values mean a better-than-average score. The exposure at time s is

$$x(s) = g(s)\, \bar{a}(s), \qquad g(s) = 1 - e^{-s/\tau}, \qquad \bar{a}(s) = \frac{1}{s} \int_0^s \tilde{a}(u)\, du, \tag{2}$$

with $\tau = \infty$ denoting the linear onset $g(s) = s/10$. Exposure, and hence the diet term, is zero at baseline.

Symmetric Choquet aggregation. A capacity on $N = \{1, \ldots, d\}$ is a set function μ with $\mu(\emptyset) = 0$, $\mu(N) = 1$, and $\mu(A) \leq \mu(B)$ for $A \subseteq B$. A 2-additive capacity is $\mu(A) = \sum_{i \in A} m_i +$

$\sum_{\{i,j\}\subseteq A} m_{ij}$, and its Choquet integral for $x \geq 0$ is $C_\mu(x) = \sum_i m_i x_i + \sum_{i<j} m_{ij} \min(x_i, x_j)$ (Grabisch, 1997). Diet exposures have both signs, so we use the symmetric integral $S = C_\mu(x^+) - C_\mu(x^-)$ with $x^\pm = \max(\pm x, 0)$ (Šipoš, 1979). Collecting the masses in $m \in \mathbb{R}^{66}$ and writing

$$\varphi(x) = \big(x_1, \ldots, x_d,\ (\psi(x_i, x_j))_{i<j}\big), \qquad \psi(x_i, x_j) = \min(x_i^+, x_j^+) - \min(x_i^-, x_j^-), \quad (3)$$

gives $S(x) = m^\top \varphi(x)$, and since $g \geq 0$, $\beta_c S_c(x(s)) = g(s)\, \theta_c^\top \varphi(\bar{a}(s))$ with $\theta_c = \beta_c m^c$. The diet term is therefore linear in the 66 numbers θ_c; the next result says which θ_c are valid.

Proposition 1. *Let $K = \{\theta \in \mathbb{R}^{66} : \theta_i \geq \sum_{j\neq i} \max(-\theta_{ij}, 0)$ for $i = 1, \ldots, d\}$. (a) $\theta = \beta m$ for some $\beta \geq 0$ and some 2-additive capacity m if and only if $\theta \in K$; K is a closed convex polyhedral cone, and every nonzero $\theta \in K$ has $\beta = \mathbf{1}^\top \theta > 0$ and $m = \theta/\beta$. Hence the diet terms of all 2-additive capacities and all effect sizes are exactly $\{g\, \theta^\top \varphi : \theta \in K \cup (-K)\}$. (b) For every capacity, $S(-x) = -S(x)$, $S(tx) = tS(x)$ for $t \geq 0$, and $|S(x)| \leq \|x\|_\infty$. The Shapley value of component i is $m_i + \frac{1}{2}\sum_j m_{ij}$ and the interaction index of pair $\{i, j\}$ is m_{ij}.*

The proof is in Appendix A. The key step is that the smallest marginal contribution of component i is $m_i - \sum_j \max(-m_{ij}, 0)$, so monotonicity is d linear constraints once each pair weight is split into its positive and negative parts, instead of $d2^{d-1}$ for a general capacity.

Exact decomposition. By Eq. 3, the diet term of outcome c is $\sum_i g\, \theta_{c,i} \bar{a}_i + \sum_{i<j} g\, \theta_{c,ij}\, \psi(\bar{a}_i, \bar{a}_j)$: 11 component parts and 55 pair parts. The pair function ψ is nonzero only when both components are on the same side of the cohort mean: a positive mass m_{ij} means that two components better (or worse) than average push further together than either alone, and a negative one that they overlap. Averaging models averages the parts, so a seed average keeps the decomposition.

3.3 Estimation

Diet quality is correlated with context (here context explains 6% of the variance of the mean diet score), so fitting D first would let it absorb the change that goes with diet. We instead partial out the context from both sides (Robinson, 1988). Taking the expectation of Eq. 1 given w_s and subtracting gives

$$\Delta_c(s) - \ell_c(w_s) = \theta_c^\top R(s) + \varepsilon_c, \qquad R(s) = g(s)\big(\varphi(\bar{a}(s)) - e(w_s)\big), \quad (4)$$

with $\ell_c(w) = \mathbb{E}[\Delta_c \mid w]$ and $e(w) = \mathbb{E}[\varphi(\bar{a}) \mid w]$. The trajectory has dropped out, and both nuisances are regressions on context alone.

Cross-fitted nuisances. The training participants are dealt into $K = 5$ folds. For each fold, an outcome network $\ell^{(k)}$ (used as $\ell(w_s) - \ell(w_0)$, zero at baseline) and an exposure network $e^{(k)}$ (66 outputs) are trained on the other folds, with the epoch chosen on the validation participants; their out-of-fold predictions give residuals $r_c = \Delta_c - \hat{\ell}_c$ and regressors $\hat{R}$ (Chernozhukov et al., 2018).

Exact fit of the capacity. For each outcome,

$$\hat{\theta}_c = \underset{\theta \in K \cup (-K)}{\arg\min}\ \frac{1}{n}\sum_n \big(r_{c,n} - \theta^\top \hat{R}_n\big)^2 + \rho \sum_{i<j} |\theta_{ij}| + \kappa \left\|\theta - \bar{\theta}\, u\right\|^2, \qquad \bar{\theta} = u^\top \theta / d, \quad (5)$$

where $u = (\mathbf{1}_d, \mathbf{0}_{55})$ is the direction of the total diet score. Splitting $\theta_{ij} = p_{ij} - q_{ij}$ with $p, q \geq 0$ makes the cone and the L_1 penalty linear, so Eq. 5 is two convex quadratic programs, one per cone, which we solve to high accuracy (Stellato et al., 2020). The *anchor* κ sets how far the weights may depart from the total score: $\kappa = 0$ fits all 66 weights, $\kappa = \infty$ gives one coefficient on the total score, and in between component weights are pulled toward equal and pair weights toward zero, without shrinking the effect along the total score. The loss is Neyman-orthogonal, so small errors in $\hat{\ell}$ and $\hat{e}$ affect $\hat{\theta}$ only at second order (Chernozhukov et al., 2018).

Choice of settings. Diet timing, onset $\tau \in \{1, 3, 10, \infty\}$ years, $\rho \in \{0, 10^{-3}\}$, and $\kappa \in \{0, 0.01, \ldots, 10, \infty\}$ are chosen jointly by 5-fold cross-validation of the orthogonal loss on the training participants, pooled over outcomes in raw units.

Forecast. With $\bar{\ell}$ and $\bar{e}$ the averages of the fold networks,

$$\hat{y}_c(s) = y_c(0) + \sigma_c \Big[\underbrace{\bar{\ell}_c(w_s) - g(s)\,\hat{\theta}_c^\top\,\bar{e}(w_s)}_{\hat{D}_c(w_s)} + g(s)\,\hat{\theta}_c^\top\,\varphi(\bar{a}(s)) \Big], \tag{6}$$

which is Eq. 1 with the diet-free trajectory $\hat{D}_c$, the forecast for the same person eating the cohort-average diet ($\bar{a} = 0$). The final forecaster averages five seeds. Whole-cohort effects get 95% intervals by resampling participants (Efron and Tibshirani, 1993) and a participant-clustered Wald test (Liang and Zeger, 1986) (Appendix B).

3.4 From predicted measurements to 3D bodies

We use TRELLIS.2 (Xiang et al., 2025a), a frozen generator that produces a textured mesh from an image in three flow-matching stages (sparse structure, shape, texture), each conditioned on self-supervised image features (Siméoni et al., 2025). Its conditioning c has 5,253,120 values. To learn how body measurements relate to c without scans, we build a reference set of 320 generated bodies. Age, sex, height, weight, and waist are resampled from cohort records (ages 17–92). A text-to-image model (Wu et al., 2025) renders a full-body photograph from a prompt stating these attributes, the background is removed (Zheng et al., 2024), and the generator's encoder produces c. We fit $P = 128$ principal directions V on the reference set and a ridge map W from seven standardized body features b^{body} (age, height, weight, waist, BMI, waist-to-height ratio, sex) to the principal scores. In use, the person's initial image is encoded once to give c_0, and at visit k the predicted height, weight, and waist from Eq. 6, with age advanced, give b_k^{body} and

$$c_k = c_0 + V\,W^\top \big(b_k^{\text{body}} - b_0^{\text{body}}\big). \tag{7}$$

The person's own encoding replaces the ridge intercept, so only the predicted change moves the conditioning. The structure and shape stages receive c_k and the texture stage receives c_0, which keeps clothes, hair, and skin unchanged; the sampling seed is the same at every visit. Height changes little (median 0.8 cm over six years) and the generator normalizes it, so we use height-relative measures.

4 Experimental Setup

Cohort. We use a controlled-access longitudinal table from two studies, with five multiple imputations of diet (Rubin, 1987). Study A enrols men and women aged 17–35 at baseline, with follow-up at +8 and +21 years that measures HOMA-IR. Study B enrols women aged 49–79, with follow-up at +1, +3, and +6 years that measures the other nine outcomes. Diet is recorded at every visit. Outcomes are observed, not imputed, and missing values are masked. After requiring baseline height, weight, and waist and at least one measured follow-up, 7,876 participants remain (3,747 in Study A, 4,129 in Study B), giving 39,380 person-imputation records (details in Appendix D).

Protocol. Participants are split 70/15/15 into training, validation, and test sets (5,514/1,181/1,181) with all records of a person in the same set. We repeat this for five participant splits and train five seeds per split, and every model and baseline is scored on the same five test sets. For forecasts of specific people and for whole-cohort effect estimates we cross-fit: participants are dealt into five folds, and each person is forecast by the models for which their fold was the test set.

Metrics. For outcome c, skill is $1 - \mathrm{MSE}_c/\mathrm{MSE}_c^{\mathrm{CF}}$, where MSE is the mean squared error and CF is the carry-forward forecast, which predicts no change from baseline (Murphy, 1988). Overall skill is $1 - \sum_c \mathrm{MSE}_c / \sum_c \mathrm{MSE}_c^{\mathrm{CF}}$ in raw units, which weights outcomes with large raw variance most; we also report the mean of the ten outcome skills and the normalized mean absolute error (nMAE), the mean of $|y - \hat{y}|/\sigma_c$. The baseline visit is never scored.

Baselines and ablations. Besides carry-forward, we use a per-outcome linear trend; ridge regression ($\alpha = 10$) of change on context, baseline outcomes, elapsed time and its square root, and the diet at baseline and at the target visit (total score, all 11 components, or baseline diet only); and histogram gradient boosting (Pedregosa et al., 2011), default and tuned. We also train a *Choquet pace model*, in which each component drives a monotone exposure clock and a general capacity combines the clocks,

Table 1: Held-out test skill relative to carry-forward (higher is better), mean over five participant splits (1,181 test participants each). "Mean" is the mean of the ten outcome skills; nMAE is lower-is-better. "Diet" is the diet input. Best value per column in bold.

Model	Diet	Overall	Mean	nMAE	Height	Weight	Waist	Syst.	Diast.	Gluc.	Insulin	HDL	LDL	HOMA
Carry-forward	–	.000	.000	.612	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Linear in time	–	.034	.028	.605	.016	.000	.006	.002	.064	.000	.004	.001	.069	.117
Ridge	total	.227	.173	.573	.052	.042	.224	.260	**.281**	.184	**.190**	.098	.254	.143
Ridge	11 comp.	.223	.170	.576	.049	.038	.222	.258	.277	.181	.186	.096	.249	.140
Ridge	baseline	.224	.171	.575	.050	.038	.220	.259	.278	.181	**.190**	.096	.250	.143
Boosting	11 comp.	.196	.112	.584	−.083	.088	.213	.236	.258	.135	−.021	.064	.241	−.015
Boosting, tuned	11 comp.	.197	.105	.587	−.088	.068	.218	.233	.253	.144	.025	.065	.241	−.111
Choquet pace model	11 comp.	.235	.169	.562	.045	.087	.177	.264	.271	.205	.135	.098	.272	.134
SCT, single model	11 comp.	.248	.198	**.561**	.077	.165	.227	.264	.276	.216	.186	**.105**	.277	.186
SCT, 5-seed average	11 comp.	**.249**	**.199**	**.561**	**.078**	**.167**	**.228**	**.265**	.277	**.217**	.187	**.105**	**.278**	**.189**

Table 2: Ablations: test skill of single models and 5-seed averages (mean over five splits), and predicted weight change for a diet one standard deviation better (mean ± SD over 25 runs; runs in which it is negative).

Variant	Single	Average	Weight change (kg)	Runs lower
SCT	.2482	.2490	-1.05 ± 0.34	25/25
start timing only	.2481	.2488	-0.93 ± 0.31	25/25
total diet score only ($\kappa = \infty$)	.2484	.2491	-1.05 ± 0.35	25/25
no anchor ($\kappa = 0$)	.2468	.2477	-1.44 ± 0.62	25/25
no anchor, no pair terms	.2468	.2477	-1.43 ± 0.61	25/25
no capacity constraint	.2420	.2429	$+0.31 \pm 0.68$	5/25
no partialling-out of diet features	.2471	.2478	-1.41 ± 0.65	25/25
no diet term	.2485	.2491	0	–
Diet term trained by gradient descent	.2451	.2485	-0.16 ± 0.22	10/25

so that diet sets only the speed of change toward a diet-free end point (25 runs; Appendix H). Ablations of SCT reuse its nuisance networks and change one part of the fit (Table 2; "no partialling-out" sets $\hat{e} = 0$). A last ablation trains the diet term by gradient descent (Appendix C).

3D evaluation. The bridge is evaluated on five 224/48/48 splits of the reference set, with all choices made on training bodies, and on a grouped split in which no test body derives from a forecaster validation or test participant. For the edit study we take 24 reference bodies balanced by sex, age third, and BMI rank, and generate each edit (± 8 kg, $+8$ cm waist, $+10$ years) with two generator seeds. On each mesh we measure a volume index (volume over height cubed) and a waist index (torso circumference at 55–65% of height, over height). At fixed density, ± 8 kg should change volume by $\pm 8/W$ for body weight W, and the change between two seeds of the unedited body is the noise level.

5 Results

5.1 Forecast accuracy

Table 1 compares SCT with the baselines. The seed-averaged forecaster has overall skill 0.249 and mean outcome skill 0.199, against 0.227 and 0.173 for the best baseline (ridge regression with the total diet score). It is ahead of that baseline and of gradient boosting on every split (Appendix E). It has the best skill in eight of ten outcomes and the lowest nMAE, and the largest gain is in weight, which shapes the 3D body (0.167 against at most 0.088). The Choquet pace model reaches 0.235, with lower weight and waist skill.

What each part of the fit does. Table 2 changes one part at a time. The accuracy comes from the trajectory: every variant with a valid capacity is within 0.002 of the model without diet; the parts of the fit matter for the diet term. Without the anchor, the 66 weights cost about 0.0013 of skill and the weight effect varies twice as much between runs; the anchor removes both. Without the capacity

Table 3: Predicted change for a diet one standard deviation better in every component at every visit. Split runs: mean over 25 runs and the number with the same sign. Whole cohort: all 7,876 participants cross-fitted, $\kappa = 0.3$ chosen by cross-validation, 95% bootstrap intervals; total score: $\kappa = \infty$, with the p-value of its clustered Wald test. Height is a control.

	Split runs		Whole cohort		
Outcome	Mean	Same sign	SCT	Total score	p
Weight (kg)	-1.05	25	-1.23 [-1.85, -0.78]	-1.34 [-1.94, -0.74]	$< 10^{-4}$
Waist (cm)	-1.19	25	-1.20 [-1.78, -0.82]	-1.40 [-2.04, -0.77]	$< 10^{-4}$
Systolic (mmHg)	-1.18	25	-1.44 [-2.74, $+0.83$]	-1.25 [-2.64, $+0.02$]	0.050
Diastolic (mmHg)	-0.75	25	-0.92 [-1.65, $+0.56$]	-0.76 [-1.51, -0.05]	0.029
Insulin (μU/mL)	-0.75	25	-0.95 [-1.82, -0.45]	-1.02 [-1.85, -0.31]	0.006
HOMA-IR	-0.41	25	-0.42 [-0.74, -0.27]	-0.36 [-0.63, -0.13]	0.007
LDL (mg/dL)	-1.21	20	-1.83 [-3.95, $+3.27$]	-1.39 [-3.59, $+0.81$]	0.24
HDL (mg/dL)	$+0.26$	18	$+0.75$ [-1.35, $+1.50$]	$+0.08$ [-0.85, $+0.88$]	0.86
Glucose (mg/dL)	$+0.53$	14	$+1.96$ [-2.32, $+3.61$]	-0.03 [-1.92, $+1.75$]	0.97
Height (cm)	$+0.12$	22	$+0.15$ [-0.36, $+0.40$]	$+0.03$ [-0.25, $+0.34$]	0.85

constraint, least squares loses 0.006 of skill and predicts that a better diet raises weight in 20 of 25 runs, so the cone is what keeps the weights meaningful. Partialling out the diet features changes little here, since context predicts diet only weakly. Trained by gradient descent after the trajectory with a context-dependent capacity, the diet term is kept in only 10 of 25 runs with about a sixth of the effect; gradient descent on the same objective keeps about 80% (Appendix C), so most of this loss comes from the other changes.

5.2 The diet term

How much diet enters. Switching the diet term off changes test skill by -0.0002 ± 0.0003, and the diet term accounts for 2.0% of the variance of predicted change in weight, 4.5% in HOMA-IR, and less in the other outcomes. No exposure form we tried improves out-of-fold fit (Appendix F). On a synthetic cohort with the same design and known diet effects (Appendix G), the diet term raises test skill by 0.042 in every one of 25 runs, so the method uses diet when it drives change.

Direction and size. In all 25 split runs (Table 3), a better diet predicts lower weight, waist, systolic and diastolic pressure, insulin, and HOMA-IR, and a worse diet the mirror image. Over the whole cohort, a better diet predicts 1.23 kg lower weight (95% interval 0.78 to 1.85), a 1.20 cm smaller waist, and lower insulin and HOMA-IR, with intervals that exclude zero; with the total score alone, diastolic pressure is also lower ($p = 0.03$). LDL, HDL, glucose, and height, which diet should not move, are not distinguishable from zero. The weight effect hardly depends on the anchor (-1.19 to -1.34 kg for κ from 0.3 to ∞).

Components and pairs. On this cohort cross-validation keeps the weights close to the total score (anchor at least 0.3 in 24 of 25 runs). A clustered Wald test of equal component weights (10 degrees of freedom; Appendix B) gives $p = 0.03$ for weight, $p = 0.006$ for glucose, and $p \geq 0.06$ for the other outcomes (waist 0.25); no outcome passes a Bonferroni correction for ten outcomes. Descriptively, for weight, trans fat has the largest Shapley value (0.168), followed by fruit (0.125) and nuts and legumes (0.103), against a uniform $1/11 \approx 0.091$, and pair masses are small (mean $|m_{ij}| = 0.005$). On the synthetic cohort, where effects differ by component, the same test rejects equal weights for all eight outcomes with a planted effect ($p \leq 0.008$) and not for height ($p = 0.58$), cross-validation chooses a weak anchor ($\kappa \leq 0.1$) in all runs, and predicted single-component effects correlate 0.73–0.91 with the true ones.

5.3 3D bodies

From measurements to conditioning. Table 4 evaluates the bridge. In the principal space, the seven body features explain $R^2 = 0.576 \pm 0.029$ of held-out variance, against 0.369 for age and sex alone and 0.183 for the nearest reference body. Removing age costs 0.091, removing weight and BMI

Table 4: Body features to generator conditioning, five splits of the reference set (224/48/48 bodies). R^2 and centred cosine on test bodies, mean $\pm$ SD over splits. The upper block is in the 128-dimensional principal space; the lower block is in the full conditioning space.

Predictor	R^2	Cosine
Nearest reference body	0.183 ± 0.093	0.607 ± 0.038
Age and sex	0.369 ± 0.035	0.599 ± 0.035
All seven features (ridge)	$\mathbf{0.576 \pm 0.029}$	$\mathbf{0.758 \pm 0.027}$
without age	0.485 ± 0.037	0.677 ± 0.046
without weight and BMI	0.554 ± 0.031	0.735 ± 0.033
without waist and waist/height	0.570 ± 0.034	0.755 ± 0.029
Ridge, decoded to full conditioning	0.471 ± 0.030	0.687 ± 0.028
Principal reconstruction (upper bound)	0.817 ± 0.013	0.901 ± 0.007

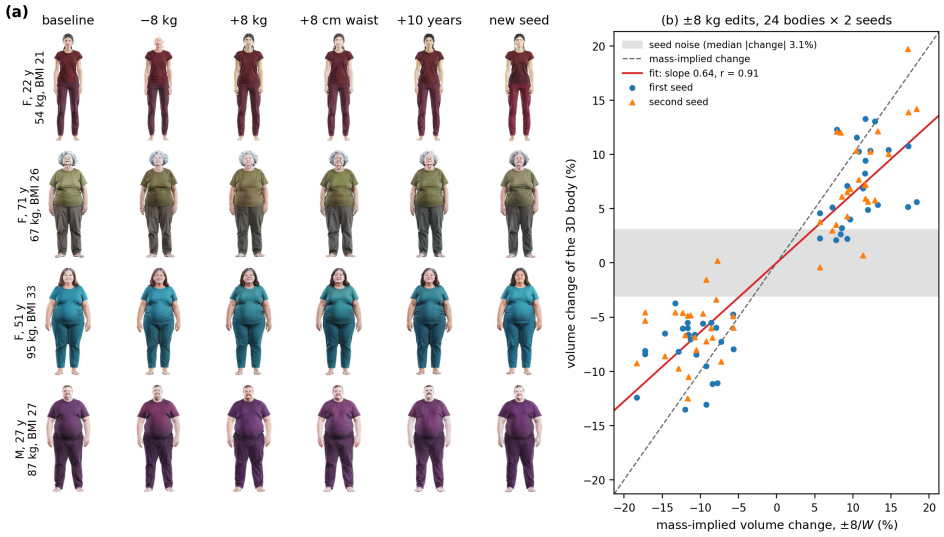

Figure 2: Controlled edits. (a) Front renders of four of the 24 bodies: baseline, single-measurement edits decoded with the baseline seed, and the unedited body with a second seed. Texture comes from the unedited conditioning, so clothes and hair stay the same. (b) Volume change of the generated body against the change implied by mass for every ± 8 kg edit (24 bodies, two seeds); the grey band is the generator's seed-to-seed change.

costs 0.022, and removing waist costs 0.006. On the grouped split (Section 4) it reaches $R^2 = 0.548$ on 38 test bodies. In the full conditioning space the map explains 0.471 of held-out variance, against an upper bound of 0.817 for a perfect prediction of the 128 principal scores.

Controlled edits. Adding 8 kg (Figure 2) grows the generated body in 47 of 48 body-seed pairs (mean volume $+7.5\%$) and removing 8 kg shrinks it in 47 of 48 (-7.1%); for 23 of 24 bodies both seeds agree on the direction. Across the 96 weight edits, the volume change is 0.64 of the mass-implied change with correlation 0.91, and the typical edit is 2.2 times the seed-to-seed change (median 3.1%). The $+8$ cm waist edit raises the waist index in 38 of 48 cases (mean $+0.9\%$), and the $+10$-year age edit increases volume in 40 of 48 (mean $+3.1\%$).

Forecast bodies. For 16 Study B women with measured six-year follow-up, chosen by baseline BMI and age, we generate bodies at enrolment, from measured follow-up values, and from the cross-fitted forecast of SCT (Appendix I, Figure 3). Bodies from measured values track the measured weight change (volume slope 0.63, correlation 0.67). The forecast is closer to the measured value than the enrolment value for 10 of 16 women in weight and 13 in waist, but its body is closer to the body from measured values than the enrolment body in 53% (volume) and 47% (waist index) of

body-seed pairs: forecast changes (standard deviation 2.5% of weight, against 6.5% measured) are below the generator's median seed-to-seed change of 3.1%.

6 Discussion and Limitations

The accuracy of SCT comes from its trajectory networks; its diet term adds an estimate of how diet relates to change, with intervals and a test against the total diet score. Because valid capacities form convex cones, the diet weights are the exact solution of a convex problem on orthogonalized residuals, so the estimate does not depend on an optimizer's path: it has the same sign in every run and recovers the overall effect on a synthetic cohort, at no cost in accuracy.

The real world cohort limits what can be learned about diet and is a common issue in this form of data. Follow-up for nine of the outcomes covers six years in older women; men contribute follow-up for HOMA-IR only, and sex is partly confounded with age band and calendar era. Attrition is substantial, and diet at Study A follow-ups is partly imputed. All relationships are observational: the effects in Table 3 are associations adjusted for context, assume that nothing unmeasured drives both diet and change, and are not estimates of what a change in diet would cause. On this cohort the diet term does not improve accuracy, and the equal-weights test gives little evidence that components matter beyond the total score, so these data do not show a benefit of the component and pair weights over the simpler total-score model. With report timing, forecasts use the diet reported at the target visit. The intervals keep the nuisance networks and settings fixed, and coverage was checked on one synthetic cohort: there the interval for the overall effect covers the truth in seven of eight outcomes, but intervals for single components reach nominal coverage only with a weak anchor (0.88 at $\kappa = 0$, 0.33 at $\kappa = 0.3$). Since cross-validation picks stronger anchors on the real cohort, the component weights and pairs reported here are exploratory; planted pairs rank highly but are recovered only in small part (Appendix G). The five splits share test participants, so their spread is descriptive.

The 3D path is generated, not measured: no participant has a scan or photograph, initial images are synthetic, and appearance is set by a prompt and seed. The waist edit reaches about a tenth of the expected change, and the age edit is about the size of the seed-to-seed change. Forecasts shrink toward the average trajectory, and changes of their size are below the generator's seed-to-seed change, so the 3D path shows edits of several kilograms but not the small changes most six-year forecasts predict. No 3D output is a likeness or an anatomical prediction for a real person.

Reproducibility Statement

The supplementary archive contains code for the forecaster, ablations, baselines, analyses, and 3D path; each step offers to generate or download the data and models it needs, and Appendix B gives all settings. The cohort is controlled-access, but a synthetic cohort with the same layout (Appendix G) runs every forecasting step. All pretrained models are public.

Ethics Statement

The health data are deidentified, used under a data use agreement, and absent from all released files. Generated bodies are synthetic, not likenesses of participants. Race and ethnicity enter the forecaster as context only and never set appearance. Since images invite stronger readings than the evidence supports, we report intervals, run-to-run agreement, and seed-to-seed variation.

AI Use Statement

Generative AI tools were used in polishing the manuscript. All experiments and results are gathered from implementation with partial code generation for prototyping and bug fixing; the authors checked the text, code, references, and numbers and are responsible for them.

REFERENCES

Rishabh Agarwal, Levi Melnick, Nicholas Frosst, Xuezhou Zhang, Ben Lengerich, Rich Caruana, and Geoffrey E. Hinton. Neural additive models: Interpretable machine learning with neural nets. In *Advances in Neural Information Processing Systems*, 2021.

Gleb Beliakov, Ana Pradera, and Tomasa Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer, Heidelberg, 2007.

Bergthor Björnsson, Carl Borrebaeck, Nils Elander, Thomas Gasslander, Danuta R. Gawel, Mika Gustafsson, Rebecka Jörnsten, Eun Jung Lee, Xinxiu Li, Sandra Lilja, et al. Digital twins to personalize medicine. *Genome Medicine*, 12:4, 2020.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Stephanie E. Chiuve, Teresa T. Fung, Eric B. Rimm, Frank B. Hu, Marjorie L. McCullough, Molin Wang, Meir J. Stampfer, and Walter C. Willett. Alternative dietary indices both strongly predict risk of chronic disease. *The Journal of Nutrition*, 142(6):1009–1018, 2012.

Gustave Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1954.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.

Michel Grabisch. The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89(3):445–456, 1996.

Michel Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92(2):167–189, 1997.

Michel Grabisch and Christophe Labreuche. Bi-capacities—I: Definition, Möbius transform and interaction. *Fuzzy Sets and Systems*, 151(2):211–236, 2005.

Michel Grabisch and Christophe Labreuche. A decade of application of the choquet and sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175(1):247–286, 2010.

Michel Grabisch and Marc Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *International Journal of Game Theory*, 28(4):547–565, 1999.

Michel Grabisch, Ivan Kojadinovic, and Patrick Meyer. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the kappalab R package. *European Journal of Operational Research*, 186(2):766–785, 2008. doi: 10.1016/j.ejor.2007.02.025.

Trevor Hastie and Robert Tibshirani. Generalized additive models. *Statistical Science*, 1(3):297–310, 1986.

Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. In *International Conference on Learning Representations (ICLR)*, 2024.

Evangelia Katsoulakis, Qi Wang, Huanmei Wu, Leili Shahriyari, Richard Fletcher, Jinwei Liu, Luke Achenie, Hongfang Liu, Pamela Jackson, Ying Xiao, et al. Digital twins for health: A scoping review. *npj Digital Medicine*, 7:77, 2024.

Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22, 1986.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, 34(6):248:1–248:16, 2015.

Yin Lou, Rich Caruana, Johannes Gehrke, and Giles Hooker. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 623–631, 2013.

Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 2017.

D. R. Matthews, J. P. Hosker, A. S. Rudenski, B. A. Naylor, D. F. Treacher, and R. C. Turner. Homeostasis model assessment: Insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, 28(7):412–419, 1985.

Cassidy McCarthy, Michael C. Wong, Jasmine Brown, Sophia Ramirez, Shengping Yang, Jonathan P. Bennett, John A. Shepherd, and Steven B. Heymsfield. Accurate prediction of three-dimensional humanoid avatars for anthropometric modeling. *International Journal of Obesity*, 48:1741–1747, 2024. doi: 10.1038/s41366-024-01614-3.

Dariush Mozaffarian, Tao Hao, Eric B. Rimm, Walter C. Willett, and Frank B. Hu. Changes in diet and lifestyle and long-term weight gain in women and men. *New England Journal of Medicine*, 364(25):2392–2404, 2011.

Allan H. Murphy. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116(12):2417–2424, 1988.

Xinkun Nie and Stefan Wager. Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika*, 108(2):299–319, 2021.

Fabian Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Peter M. Robinson. Root-N-consistent semiparametric regression. *Econometrica*, 56(4):931–954, 1988.

Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York, 1987.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. DeepSeekMath: Pushing the limits of mathematical reasoning in open language models, 2024.

Lloyd S. Shapley. A value for n-person games. In Harold W. Kuhn and Albert W. Tucker, editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton, NJ, 1953.

Oriane Siméoni et al. DINOv3, 2025.

Ján Šipoš. Integral with respect to a pre-measure. *Mathematica Slovaca*, 29(2):141–155, 1979.

Bartolomeo Stellato, Goran Banjac, Paul Goulart, Alberto Bemporad, and Stephen Boyd. OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation*, 12(4): 637–672, 2020.

Michio Sugeno. *Theory of Fuzzy Integrals and its Applications*. PhD thesis, Tokyo Institute of Technology, 1974.

Ali Fallah Tehrani, Weiwei Cheng, Krzysztof Dembczyński, and Eyke Hüllermeier. Learning monotone nonlinear models using the Choquet integral. *Machine Learning*, 89(1–2):183–211, 2012.

Dmitry Tochilkin, David Pankratz, Zexiang Liu, Zixuan Huang, Adam Letts, Yangguang Li, Ding Liang, Christian Laforte, Varun Jampani, and Yan-Pei Cao. Triposr: Fast 3d object reconstruction from a single image, 2024.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992.

Chenfei Wu et al. Qwen-image technical report, 2025.

Jianfeng Xiang, Xiaoxue Chen, Sicheng Xu, Ruicheng Wang, Zelong Lv, Yu Deng, Hongyuan Zhu, Yue Dong, Hao Zhao, Nicholas Jing Yuan, and Jiaolong Yang. Native and compact structured latents for 3D generation, 2025a. URL https://arxiv.org/abs/2512.14692. Tech report.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3D latents for scalable and versatile 3D generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21469–21480, 2025b. doi: 10.1109/CVPR52734.2025.02000.

Peng Zheng, Dehong Gao, Deng-Ping Fan, Li Liu, Jorma Laaksonen, Wanli Ouyang, and Nicu Sebe. Bilateral reference for high-resolution dichotomous image segmentation. *CAAI Artificial Intelligence Research*, 3:9150038, 2024.

A PROOFS

Linearity. For $x \geq 0$ the 2-additive Choquet integral is $C_\mu(x) = \sum_i m_i x_i + \sum_{i<j} m_{ij} \min(x_i, x_j)$ (Grabisch, 1997). Applying it to x^+ and x^- and subtracting gives $S(x) = \sum_i m_i(x_i^+ - x_i^-) + \sum_{i<j} m_{ij}\psi(x_i, x_j) = m^\top \varphi(x)$, using $x_i^+ - x_i^- = x_i$. Since $\psi(tx_i, tx_j) = t\psi(x_i, x_j)$ for $t \geq 0$, $\varphi(g\bar{a}) = g\varphi(\bar{a})$ and $\beta S(g\bar{a}) = g\,(\beta m)^\top \varphi(\bar{a})$. If $x_i, x_j > 0$ then $\psi = \min(x_i, x_j)$; if $x_i, x_j < 0$ then $\psi = \max(x_i, x_j)$, the one closer to zero; otherwise $\psi = 0$.

Proposition 1(a). For a 2-additive set function with masses θ, the marginal contribution of i to $A \not\ni i$ is $\theta_i + \sum_{j \in A} \theta_{ij}$, and its minimum over A is $\theta_i - \sum_{j \neq i} \max(-\theta_{ij}, 0)$, attained at $A = \{j : \theta_{ij} < 0\}$. Monotonicity, which is equivalent to all marginal contributions being nonnegative, therefore holds exactly when $\theta \in K$. If $\theta = \beta m$ with $\beta \geq 0$ and m a capacity, the marginal contributions of θ are β times those of m, so $\theta \in K$. Conversely, let $\theta \in K$ be nonzero and write $\theta(A)$ for its set function. $\theta(N) = \mathbf{1}^\top \theta$ is a sum of marginal contributions along any ordering of N, so it is nonnegative. If it were zero, every marginal contribution along every ordering would be zero; taking each i first gives $\theta_i = 0$, then $\theta \in K$ forces $\theta_{ij} \geq 0$, and $\theta(N) = \sum \theta_{ij} = 0$ gives $\theta_{ij} = 0$, a contradiction. So $\beta = \theta(N) > 0$, and $m = \theta/\beta$ is monotone with $m(\emptyset) = 0$ and $m(N) = 1$. K is the projection onto θ of the polyhedron $\{(\theta, p, q) : \theta_{ij} = p_{ij} - q_{ij},\ p, q \geq 0,\ \theta_i \geq \sum_j q_{ij}\}$, hence a closed convex polyhedral cone; the projection is exact because $q_{ij} = \max(-\theta_{ij}, 0)$ is feasible whenever any q is. Negative effect sizes give $-K$. □

Proposition 1(b). C_μ is positively homogeneous and, for $x \geq 0$, satisfies $0 \leq C_\mu(x) \leq \max_i x_i$ because it is a Choquet integral with respect to a capacity. Since $(-x)^+ = x^-$, $S(-x) = C_\mu(x^-) - C_\mu(x^+) = -S(x)$, and homogeneity carries over. Both terms of S lie in $[0, \|x\|_\infty]$, so $|S(x)| \leq \|x\|_\infty$. The Shapley value of a 2-additive capacity is $m_i + \frac{1}{2}\sum_j m_{ij}$ and its pairwise interaction index is m_{ij} (Grabisch, 1997).

The partialled-out equation and orthogonality. Taking $\mathbb{E}[\cdot \mid w_s]$ of Eq. 1 gives $\ell_c(w_s) = D_c(w_s) + g(s)\theta_c^\top e(w_s)$, because g is a function of s, which is part of w_s. Subtracting from Eq. 1 gives Eq. 4. The derivative of $\mathbb{E}[(\Delta - \ell - \theta^\top g(\varphi - e))^2]$ with respect to ℓ in any direction $h(w)$ is $-2\,\mathbb{E}[\varepsilon\, h(w)] = 0$, and with respect to e in direction $h(w)$ it is $2\,\mathbb{E}[\varepsilon\, g\, \theta^\top h(w)] = 0$ at the true (ℓ, e, θ), so the loss is insensitive to first-order errors in both nuisances (Chernozhukov et al., 2018). The constraint set and the penalties do not involve the nuisances.

Decomposition of averages. For an average of M models, each forecast is $\frac{1}{M}\sum_M [\hat{D}^{(M)} + \sum_i P_i^{(M)} + \sum_{i<j} P_{ij}^{(M)}]$, so the averaged parts again add up exactly. The average of $\hat{\theta}$ over seeds lies in K or $-K$ whenever all seeds agree on the sign of the effect, which they do for the six outcomes with a clear effect.

B Estimation Details

Networks. Outcome and exposure networks take $(z, b, \log(1+s), s/10)$ and have two hidden layers of 96 SiLU units with dropout 0.15. The outcome network has one output per outcome and is trained on the masked mean squared error of the scaled change $\ell(w_s) - \ell(w_0)$; the exposure network predicts the 66 standardized features $\varphi(\bar{a}(s))$, with one network per diet timing. Both use AdamW (learning rate 1.5×10^{-3}, weight decay 10^{-3}, cosine schedule, gradient clipping at 5) for 40 epochs with batch size 1,024 on all five imputations, keeping the epoch with the highest validation skill (outcome network) or lowest validation error (exposure network). The exposure networks explain 9% (start timing) and 6% (report timing) of the variance of the standardized features on validation participants.

Quadratic programs. For each outcome and setting, Eq. 5 is solved over K and over $-K$ with variables $(\theta_{1:d}, p, q)$, $d + 110$ in all, and constraints $\theta_i - \sum_j q_{ij} \geq 0$, $p, q \geq 0$; the better of the two solutions is kept. We use OSQP with absolute and relative tolerances of 10^{-10} and solution polishing (Stellato et al., 2020), and move the result into the cone by raising singletons where the solver tolerance leaves a violation below 10^{-9}. With no pair terms, the program has only the d singleton weights, constrained to one sign. The total-score fit ($\kappa = \infty$) is one least-squares coefficient on $\sum_i R_i$. One run, with both timings, all settings, and all ablations, takes 5 to 20 minutes, mostly the programs on the CPU; the networks take about two minutes on one RTX A6000 GPU.

Settings chosen. The anchor grid is $\{0, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, \infty\}$; each candidate is solved on four folds of the training participants and scored on the fifth, and validation participants serve only for early stopping. Across the 25 split runs, cross-validation chose the report timing in 19, the onset $\tau = 1$ year in 17, and the anchor $\kappa = 1$ (8 runs), 3 (7), 0.3 (4), ∞ (3), 10 (2), and 0.1 (1); the cross-validated loss is flat near its optimum, so single runs differ. On the whole cohort it chose the report timing, $\tau = 1$, $\rho = 10^{-3}$, and $\kappa = 0.3$, the setting chosen most often by the split runs. The cross-validated gain over no diet at that setting is small and peaks at $\kappa = 0.3$ (6.7×10^{-4}, against 4.1×10^{-4} at $\kappa = 0$ and 2.1×10^{-4} at $\kappa = \infty$). On the synthetic cohort it chose the start timing, $\tau = 3$, and $\kappa = 0.1$ (17 runs) or 0.03 (8).

Intervals and tests. Whole-cohort estimates use the cross-fitted nuisances of all participants, 99,405 visit rows. Bootstrap intervals resample participants with all their records ($B = 500$), re-solve Eq. 5 at the chosen setting, and report the 2.5% and 97.5% quantiles of the predicted change. The total-score test uses a participant-clustered sandwich variance for the single coefficient. For the component tests, the 11 component weights $\hat{\theta}$ are fitted by least squares without constraints or pairs, with clustered variance $\hat{V}$. The equal-weights test has null $\theta_1 = \cdots = \theta_{11}$, written $C\theta = 0$ with $C = [I_{10} \; -\mathbf{1}]$, and statistic $(C\hat{\theta})^\top (C\hat{V}C^\top)^{-1} C\hat{\theta}$ with 10 degrees of freedom; it asks whether separate weights add to the total score. The zero test, $\theta = 0$ with 11 degrees of freedom, asks only whether diet enters at all (Table 5).

C Diet Term Trained by Gradient Descent

This ablation uses the diet term of Eq. 1 with a per-component scale, $x_j = \lambda_j g(s) \bar{a}_j$ with learned $\lambda_j > 0$ and τ, a capacity that depends on context, and gradient training.

Parameterization. The capacity is

$$\tilde{m}_{ij} = \tanh(u_{ij}), \qquad \tilde{m}_i = \textstyle\sum_{j \neq i} \max(-\tilde{m}_{ij}, 0) + \operatorname{softplus}(r_i), \qquad m = \tilde{m}/\tilde{\mu}(N), \quad (8)$$

with $\tilde{\mu}(N) = \sum_i \tilde{m}_i + \sum_{i<j} \tilde{m}_{ij}$, and context shifts the raw parameters through a low-rank code, $r_i(z) = r_i + \langle L_i, h(z) \rangle$ and $u_{ij}(z) = u_{ij} + \langle L_{ij}, h(z) \rangle$ with a learned code $h(z) \in \mathbb{R}^4$, per outcome. For every (r, u) this is a capacity: the marginal contribution of i to any A before normalization is $\tilde{m}_i + \sum_{j \in A} \tilde{m}_{ij} \geq \operatorname{softplus}(r_i) > 0$, and every 2-additive capacity with positive marginal contributions is reached (take $t > 0$ small, $u_{ij} = \operatorname{artanh}(t m_{ij})$, $r_i = \operatorname{softplus}^{-1}(t(m_i - \sum_j \max(-m_{ij}, 0)))$). The effect size $\beta_c(z, b)$ is linear in context and starts at zero, and the trajectory is $D_c = f_c(z, b, s) - f_c(z, b, 0)$ for a network f_c of the same size as the outcome networks.

Table 5: Participant-clustered Wald tests of the 11 component weights, fitted without constraints on the cross-fitted rows at the chosen setting. Zero: all weights zero (11 degrees of freedom). Equal: all weights equal, i.e. the total score suffices (10 degrees of freedom). Synthetic: the synthetic cohort at its chosen setting, where every outcome but height has planted effects that differ by component.

	Real cohort				Synthetic	
Outcome	Zero χ^2_{11}	p	Equal χ^2_{10}	p	Equal χ^2_{10}	p
Weight	40.5	$< 10^{-4}$	20.4	0.03	37.3	$< 10^{-4}$
Waist	30.6	0.001	12.5	0.25	48.0	$< 10^{-4}$
Systolic	14.5	0.21	11.2	0.34	23.9	0.008
Diastolic	19.3	0.06	14.6	0.15	35.3	10^{-4}
Insulin	14.1	0.22	4.1	0.94	35.0	10^{-4}
HOMA-IR	19.4	0.05	17.7	0.06	65.3	$< 10^{-4}$
LDL	15.9	0.15	14.6	0.15	79.4	$< 10^{-4}$
HDL	14.7	0.20	14.7	0.14	32.7	0.0003
Glucose	24.7	0.01	24.6	0.006	41.4	$< 10^{-4}$
Height	13.4	0.27	13.4	0.20	8.5	0.58

Training. Stage 1 fits the trajectory alone ($\beta \equiv 0$) by the mean squared scaled residual for 40 epochs, keeping the epoch with the highest validation skill. Stage 2 freezes the trajectory and trains the capacity, β, λ, and τ for 20 epochs at learning rate 5×10^{-4}, with an L_1 penalty of weight 10^{-3} on the pair masses and their context loadings. The stage-1 model competes in stage 2, so the diet term is kept only if it improves validation skill. Table 6 lists the settings we tried; choices were made on validation skill.

Table 6: Gradient-trained diet term: training variants, mean over five splits. Rows with five seeds give the mean of the single models, and each indented "5-seed average" row gives the skill of their average. The single-seed rows change one setting of the chosen configuration.

Variant	Seeds	Val. skill	Test skill	Test outcome mean
Squared error, epoch by validation skill (chosen)	5	.2703	.2451	.1954
5-seed average	5	.2732	.2485	.2008
Huber-core Gaussian likelihood, epoch by validation skill	5	.2706	.2436	.1865
5-seed average	5	.2731	.2466	.1902
Likelihood loss, epoch by validation nMAE	3	.2679	.2384	.1743
Dropout 0.3	1	.2689	.2475	.1952
60 epochs	1	.2699	.2466	.1982
Width 192	1	.2698	.2420	.1967
Learning rate 3×10^{-3}	1	.2694	.2450	.1994
Weight decay 10^{-2}	1	.2699	.2472	.2012
Linear path next to the trajectory network	1	.2700	.2441	.1986
Additive capacity, 5-seed average	5	.2732	.2484	.2008
One capacity for everyone, 5-seed average	5	.2738	.2487	.2007
One capacity per side of the average, 5-seed average	5	.2732	.2484	.2009
No diet term, 5-seed average	5	.2730	.2486	.2007

Results. Validation keeps the diet term in 10 of 25 runs, and switching it off changes test skill by -0.0001 on average. In the ten runs that keep it, a diet one standard deviation better predicts -0.41 kg weight and -0.43 cm waist, with the same sign in all ten; averaged over all 25 runs the effects are -0.16 kg and -0.17 cm. Fitting the trajectory first lets it absorb the part of the change that goes with diet, and the validation choice then often removes the rest, which is why the effect is smaller than that of SCT. On the synthetic cohort it recovers 73–94% of the planted effect of a better diet in seven outcomes (Table 11). Cross-fitted models of this kind dropped the diet term in all 25 fold runs.

Same objective, gradient descent. This ablation changes the capacity, the training order, and the selection rule together with the optimizer. To isolate the optimizer, we fit Eq. 5 at the whole-cohort setting on the same cross-fitted rows, with the same features, cone, pair penalty, and anchor, on four of the five cross-fitting folds, validating on the fifth. The cone is parameterized so that every iterate is valid (pair masses $p - q$ and singletons $\sum_j q_{ij}$ plus a positive slack, all through softplus), both signs

are tried, and Adam is used. Run to convergence (20,000 full-batch steps), gradient descent reaches the exact solution (weight effect -1.24 ± 0.16 kg, waist -1.20 ± 0.10 cm over folds, identical to three decimals). With minibatches of 1,024, learning rate 1.5×10^{-3}, and the best of 40 epochs on validation, it gives -0.98 ± 0.12 kg and -1.02 ± 0.15 cm, and the rule that keeps the diet term only if it beats no diet on validation keeps it in all five folds. Early stopping thus shrinks the effect by about a fifth; dropping the term in 15 of 25 runs and the smaller effect of the gradient-trained ablation come mainly from fitting the trajectory first with a context-dependent capacity.

One capacity per side. Replacing the symmetric integral with $S = C_{\mu^+}(x^+) - C_{\mu^-}(x^-)$, the two-capacity form of cumulative prospect theory (Tversky and Kahneman, 1992; Grabisch and Labreuche, 2005), lets diets better and worse than average have different weights. We parameterize μ^- as μ^+ with offsets that start at zero and carry an L_1 penalty. Refitted on the frozen trajectories, the two sides give mirror-image effects within 1.2% in eight outcomes (weight -0.57 and $+0.57$ kg, LDL -1.50 and $+1.51$ mg/dL), and the symmetric refit within 1.6%, so the symmetric integral matches these data. On the synthetic cohort, the per-side capacity leaves test skill at 0.214 and does not change the recovered pair effects.

D Cohort Details

Table 7: Study design after filtering.

	Study A	Study B
Participants before / after filtering	5,114 / 3,747	5,600 / 4,129
Sex	men and women	women
Baseline age	17–35	49–79
Follow-up (years)	+8, +21	+1, +3, +6
Measured at follow-up	HOMA-IR only	panel of nine outcomes
Diet (AHEI-2010 components)	every visit (partly imputed)	every visit

Weight is computed from BMI and height; the source stores pounds for one study and kilograms for the other. Measured follow-up values per outcome across all splits and imputations are: height 51,965; weight 51,830; waist 45,615; systolic 52,270; diastolic 52,255; glucose 52,520; insulin 36,005; HDL 52,300; LDL 51,075; HOMA-IR 33,345. Of 1,526,635 component scores, 6.0% fall outside $[0, 10]$ and are clamped. The test set of the first split has 621 Study B and 560 Study A participants. Over follow-up (median six years, $n = 4{,}118$), the median absolute change is 0.8 cm in height, 3.1 kg in weight, and 4.0 cm in waist; 90% of height changes are below 2.1 cm.

E Full Accuracy Results

Table 8: Overall test skill per participant split. SCT is the 5-seed average; baselines are refitted on each split.

Split	SCT	Ridge, total	Ridge, 11 comp.	Boosting	Boosting, tuned
0	**.261**	.227	.223	.185	.205
1	**.247**	.231	.226	.189	.183
2	**.208**	.194	.192	.149	.144
3	**.267**	.236	.232	.249	.246
4	**.262**	.245	.241	.206	.206
Mean	**.249**	.227	.223	.196	.197

The Choquet pace model has overall skill 0.235 ± 0.020 over its 25 runs (range 0.201–0.263). Cross-fitted SCT forecasts (five folds, five seeds) have overall skill 0.2685 pooled over all participants, against 0.2683 without the diet term and 0.2684 for the total score alone; the cross-fitting folds differ from the five random splits, so these values are not directly comparable with Table 1.

F Diet Term Details

Table 9: Diet term on the real cohort, mean over the 25 split runs. Paired gain: test skill minus that of the same model without the diet term. Effects: raw units, for a diet one standard deviation better at every visit and from the first follow-up on. Variance share: variance of the diet part over variance of the predicted change.

	Height	Weight	Waist	Syst.	Diast.	Glucose	Insulin	HDL	LDL	HOMA
+1 SD, every visit	+0.12	−1.05	−1.19	−1.18	−0.75	+0.53	−0.75	+0.26	−1.21	−0.41
SD over runs	0.12	0.34	0.19	0.33	0.20	1.24	0.30	0.39	1.22	0.14
+1 SD, after baseline	+0.11	−0.97	−1.09	−1.10	−0.70	+0.63	−0.67	+0.24	−1.03	−0.37
Variance share (%)	0.81	1.96	1.43	0.17	0.26	0.12	0.89	0.26	0.19	4.47

Table 9 gives per-outcome details. The paired gain over all outcomes is -0.0002 ± 0.0003. Without the anchor ($\kappa = 0$), the effects are larger and vary more between runs (weight -1.44 ± 0.62 kg, systolic pressure -2.11 ± 0.83 mmHg), and the whole-cohort intervals are wider (weight -1.56 $[-2.76, -1.04]$ kg, a width of 1.7 against 1.1 kg). Shapley values of the capacity, as a share of the effect and averaged over runs, lie between 0.046 and 0.168 for weight (largest: trans fat, fruit, nuts and legumes, red and processed meat), between 0.054 and 0.134 for waist (red and processed meat, trans fat, fruit), and between 0.047 and 0.199 for LDL (vegetables, whole grains, fruit). The most frequent leading pairs are fruit with trans fat for weight (18 of 25 runs) and waist (19), vegetables with whole grains for LDL (16), and sugar-sweetened drinks with red and processed meat for HOMA-IR (13).

Why diet adds no accuracy here. Context explains 6% of the variance of the mean diet score with ridge regression and 4% with gradient boosting, so partialling-out mainly removes the part of diet that the trajectory already explains. Between 36% and 63% of the variance of each component is change within a person from visit to visit. In a separate check on the out-of-fold residuals of the outcome networks, 14 forms of the diet exposure (level or change since baseline, with or without the baseline diet or the last report, additive or with pairs) all lowered the pooled out-of-fold R^2 slightly, by 0.0002 to 0.013, the forms with pairs the most. The diet signal is real but low-dimensional and small next to each outcome's own variation over six years.

G Synthetic Cohort

Design. The synthetic cohort copies the real study design and rounded summary statistics (means, standard deviations, missingness rates, sex and race mix), not records. It has 3,400 Study A and 2,600 Study B participants (4,898 after the same filters). Each diet score is a clipped Gaussian around a study-specific mean, correlated through a latent diet quality. Between visits, each outcome changes at a rate age drift + person slope + $\sum_j \beta_{cj} z_j$ + planted pairs, applied to $h(s) = s/(1 + s/6)$, where $z_j = (a_j - 5)/3.5$ are standardized diet scores on the 0–10 scale and each planted pair adds $\gamma \max(z_a, 0) \max(z_b, 0)$. The planted pairs are vegetables with fruit for weight ($\gamma = -0.35$) and waist (-0.45), nuts and legumes with polyunsaturated fat for LDL (-2.5), and sugar-sweetened drinks with whole grains for log glucose (-0.015). Diet does not depend on context. Measurement noise, attrition, and imputed diet follow the real design. All models are trained and scored with the same protocol as on the real cohort (five splits, five seeds).

Accuracy and diet gain. Table 10 gives test skill. SCT has the best overall skill (0.219 against 0.204–0.212 for ridge and 0.214 for the gradient-trained diet term). The diet term improves test skill by $+0.042 \pm 0.010$, positive in all 25 runs and largest for LDL ($+0.074$), weight ($+0.055$), waist ($+0.053$), and HOMA-IR ($+0.045$), and it accounts for 38–54% of the variance of predicted change for these outcomes. The total score alone reaches 0.211, so the component weights add accuracy when effects differ by component.

Recovery. Table 11 compares predicted and true changes. SCT recovers 85–111% of the planted effect of a better diet in seven outcomes and overestimates insulin, as every model does. Per person, predicted and true changes correlate 0.73–0.82 in six outcomes (0.42–0.57 for the gradient-trained

Table 10: Synthetic cohort, test skill relative to carry-forward, mean over five splits. The pace model has four runs on two splits.

Model	Overall	Outcome mean	Weight	Waist	LDL	HOMA-IR
Linear in time	.087	.058	.001	.007	.166	.243
Ridge, total	.204	.166	.061	.163	.284	.314
Ridge, 11 components	.210	.167	.068	.153	.299	.325
Ridge, baseline diet	.212	.170	.070	.154	.299	.332
Boosting	.150	.108	.024	.079	.259	.327
Choquet pace model	.192	.146	.024	.113	.278	.285
Gradient-trained diet term, 5-seed avg.	.214	.173	.086	.161	.299	**.333**
SCT, total score only, 5-seed avg.	.211	.170	.083	.165	.299	.316
SCT, no diet term, 5-seed avg.	.176	.145	.037	.113	.236	.279
SCT, single model	.218	.174	.091	.166	**.310**	.323
SCT, 5-seed average	**.219**	**.175**	**.092**	**.167**	**.310**	.323

Table 11: Recovery of planted diet effects on 367 Study B test participants per split at +6 years, 25 runs. True and predicted: mean change when every component is 1 SD (3.5 points) better at every visit. Ratio: median over runs of predicted/true. Component r: median correlation over the 11 components between predicted and true single-component effects.

	Weight	Waist	Syst.	Diast.	Glucose	Insulin	HDL	LDL
True change	−3.42	−4.22	−4.91	−2.05	−9.30	−2.64	+3.84	−21.73
SCT, predicted	−3.16	−4.25	−5.09	−2.22	−9.05	−3.96	+3.19	−19.73
SCT, ratio	0.93	1.01	0.99	1.11	0.96	1.46	0.85	0.90
SCT, component r	0.80	0.78	0.91	0.76	0.91	0.89	0.86	0.73
Gradient-trained, ratio	0.77	0.87	0.81	0.94	0.87	1.43	0.77	0.73
Gradient-trained, component r	0.70	0.75	0.85	0.68	0.93	0.84	0.83	0.70
Pace model, ratio	0.00	−0.00	0.00	0.05	−0.01	−0.01	0.01	0.08

term) and weakly for glucose and insulin (0.10, 0.18). Planted component effects have the right sign in 99–100% of cases. The total score alone recovers the overall size (ratios 0.72–1.04 in seven outcomes) but gives every component the same weight. The Choquet pace model, in which diet only sets the speed toward a diet-free end point, recovers less than 1% of the effect in six outcomes and 5–8% in the other two (four runs).

Table 12: Synthetic cohort, whole-cohort 95% bootstrap intervals (4,898 participants): share covering the true effect of a diet 1 SD better in all components (8 outcomes) and in one component at a time (88 effects, 50 of them truly zero).

Anchor	All components	One component	truly zero	nonzero
Chosen by cross-validation ($\kappa = 0.03$)	0.88	0.83	0.84	0.82
$\kappa = 0$	0.88	0.88	0.90	0.84
$\kappa = 0.3$	0.88	0.33	0.24	0.45
$\kappa \geq 1$	0.75	0.06–0.09	0.00	0.13–0.21

Intervals. Table 12 reports coverage. The interval for the overall effect covers the truth in seven of eight outcomes for every anchor up to 0.3; the miss is insulin, whose effect every model overestimates (HDL at $\kappa = 0.3$). Intervals for single components need a weak anchor: at the anchor chosen by cross-validation they cover 83% of the effects, but at $\kappa = 0.3$ only 33%, because the anchor pulls the components toward equal weights. The total-score test detects all eight planted effects ($|t|$ from 5.1 to 12.9) and not height ($t = -1.85$). Coverage here is the share of the planted effects in one synthetic cohort that the intervals contain, not coverage over repeated cohorts, and the intervals keep the nuisance networks and chosen settings fixed. The equal-weights test rejects for all eight outcomes with planted effects and not for height (Table 5).

Planted pairs. For each planted pair we compute the model's interaction effect, the change when both components improve minus the changes when each improves alone, and rank it among the 55 pairs of its outcome by size. Median ranks over 25 runs are 5 for LDL, 6 for glucose, 16 for waist,

and 22 for weight (a random pair has median rank 28). The predicted interaction effects have the right sign for weight, waist, and LDL but are small: -0.001 kg, -0.005 cm, and -0.11 mg/dL against true -0.20, -0.26, and -1.78; for glucose, $+0.03$ against -0.67 mg/dL. Three checks separate the form of the pair term from the evidence in the data. (i) On noise-free diets drawn from the generator, least squares on per-component terms explains 74–80% of the variance of the planted pair part for weight, waist, and LDL, adding the signed pair term ψ raises this to 76–80%, and one pair term per side to 95–98%. (ii) With one capacity per side (Appendix C) the recovered pair effects do not change. (iii) Regressing the scaled training residual after the trajectory on per-component terms and the two side terms of one pair, the better-side coefficient of the planted pairs for weight, waist, and LDL has the right sign in all five splits but lies only 1.4–2.6 standard errors from zero, typical of unplanted pairs. Diet components are correlated through a latent diet quality, so many pairs produce similar features, and at this cohort size the planted pairs do not stand out from them.

H Choquet Pace Model

The pace model shares the setup of Section 3.1. Each component j has a positive exposure rate $\rho_j(s)$ from a small network of $(\log(1+s), a_{k,j}, a_{0,j}, a_{k,j} - a_{0,j})$, and $q_j(s) = R_j/(1+R_j)$ with $R_j = \int_0^s \rho_j$ is a progress coordinate in $[0, 1)$. For each outcome a general capacity on all 2^{11} subsets, built in order of subset size so that every parameter value is valid, combines the coordinates by the Choquet integral $\Phi_c = C_{\mu^c}(q)$, and $y_c(s) = y_c(0) + \sigma_c A_c(z, y_0)\Phi_c(s)$, where A_c is a bounded long-run change predicted from context and baseline values. Diet therefore changes only how fast a person moves toward an end point that does not depend on diet. Training minimizes a Huber-core Gaussian likelihood with capacity regularizers and alternates with a group-relative policy-gradient search over the capacity (Williams, 1992; Shao et al., 2024); the epoch with the lowest validation nMAE is kept. Its fitted Shapley values were close to uniform (0.083–0.102 in every outcome), and it responds very little to diet changes on the synthetic cohort (Table 11).

I 3D Details

Reference set. The 320 bodies come from 315 unique source participants; the five duplicated sources are kept together in every grouped split. Each prompt describes the person's sex, age, and body build together with skin tone, hair, clothing colour, a plain studio background, and a relaxed A-pose; each body has its own seed. Images are generated with Qwen-Image-2.1, backgrounds are removed with BiRefNet, and conditioning features are computed at 512 and 1,024 pixels. The deployed principal basis is fitted on 288 bodies, with 32 held out to check reconstruction, and the deployed ridge map ($\alpha = 1$) on all 320. On the grouped split the basis captures 95.5% of training variance.

Generation. Bodies are generated with the 1,536-voxel cascade of TRELLIS.2 (4B parameters), with the sparse-structure and shape stages conditioned on the shifted condition and the texture stage on the original one; shifting the texture condition too changed hair colour and skin tone. Meshes are hole-filled; textured meshes are exported with 2,048-pixel texture maps and rendered at 1,024 pixels under a fixed studio light. One body variant takes about two minutes on one GPU.

Measurements. The generator scales each body to unit height H. We slice the mesh horizontally every 0.5% of H, fill each slice, and sum slice areas to get the volume index V/H^3. For torso slices between 40% and 80% of H, a morphological opening removes arms resting against the body and the largest region is kept; the waist index is the mean convex-hull perimeter of that region between 55% and 65% of H, divided by H.

Edits in detail. Across both seeds, $+8$ kg changes the waist index by $+4.7\%$ (47 of 48 up) and -8 kg by -4.1% (47 of 48 down); the $+8$ cm waist edit changes it by $+0.9\%$ against an expected $+9.7\%$ (slope 0.10 and 0.12 for the two seeds). Changes of the same body under the two seeds correlate weakly (0.24 for $+8$ kg volume), so per-body sizes of the response are noisy even when the direction is stable.

Forecast bodies. The 16 women are the eligible reference bodies (Study B source participant with weight and waist measured at six years) nearest to the centres of the eight cells formed by baseline-BMI quartiles and age halves, two per cell. Each forecast is the average of the five cross-fitted models whose test fold contains the woman, averaged over her five imputation records. Enrolment, measured, and forecast bodies are generated with the reference seed and again with a second seed (32 body-seed pairs). SCT's mean absolute errors over the 16 women are: height 0.61 cm (carry-forward 0.71), weight 3.76 kg (3.73), waist 4.96 cm (6.03). The forecast is closer to the measured value than carry-forward for 10 of 16 women in weight and 13 in waist; over all measured Study B follow-ups, cross-fitted forecasts beat carry-forward for weight (mean absolute error 3.82 against 3.88 kg) and waist (4.53 against 4.85 cm). The forecast body's volume change correlates 0.41 with that of the body from measured values and has the same direction in 59% of body-seed pairs; for the waist index, 0.47 and 59%. The forecast body is closer to the body from measured values than the enrolment body is in 53% of body-seed pairs for volume (median gap 5.3% against 4.5%) and 47% for the waist index (2.1% against 2.6%). The standard deviation of forecast weight change across the women is 2.5%, against 6.5% for the measured change. With the forecasts of the gradient-trained ablation, which are within 0.9 kg of those of SCT, the same comparison gives 34% and 41%, so at this sample size the fraction is noisy. Figure 3 shows all 16.



Figure 3: Forecast bodies for 16 Study B women. Each panel shows the body generated at enrolment, with the outline of the body generated from the measured values six years later (blue) and from the cross-fitted forecast for the same visit (red). Labels give enrolment age, weight, and waist, then measured (meas) and forecast (fc) weight/waist.