Supplementary Materials

Companion materials for the manuscript:

> **Moderate-to-severe olfactory dysfunction marks accelerated phenotypic aging in U.S. adults** Olcan, C. (2026). *(In submission.)*

Companion deposits (all live): - Code repository (Zenodo, MIT-licensed): https://doi.org/10.5281/zenodo.20059714 - Discovery archive (OSF, post-hoc archival): https://doi.org/10.17605/OSF.IO/7Y5FU - Pre-registered external replication protocol (OSF Registration): https://doi.org/10.17605/OSF.IO/JK2G7

Supplementary Appendix S0: STROBE-22 Reporting Checklist

This appendix maps each item of the **Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)** checklist (von Elm et al. 2007; STROBE-22 update) to the corresponding section of the manuscript "Moderate-to-severe olfactory dysfunction marks accelerated phenotypic aging in U.S. adults" (Olcan 2026).

STROBE provides a 22-item reporting framework for observational studies, including cohort, case-control, and cross-sectional designs. This study includes both cross-sectional (Aim 1) and prospective cohort (Aim 2) components; the checklist below addresses both.

Title and abstract

Item	Recommendation	Reported on
1 (a)	Indicate the study's design with a commonly used term in the title or the abstract	Title and abstract: "cross-sectional and prospective" framing throughout abstract
1 (b)	Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract (191 words; Methods, Results, and balanced "research signal not clinical adjunct" framing)

Introduction

Item	Recommendation	Reported on
2	Background/rationale: explain the scientific background and rationale for the investigation being reported	§1 Introduction (paragraphs 1–4)
3	Objectives: state specific objectives, including any pre-	§1 Introduction (paragraph 5: three pre-specified

Item	Recommendation	Reported on
	specified hypotheses	hypotheses); Methods §2.5

Methods

Item	Recommendation	Reported on
4	Study design: present key elements of study design early in the paper	§2.1 Study population (cross-sectional NHANES 2013–2014 + prospective NDI follow-up)
5	Setting: describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	§2.1: NHANES 2013–2014 cycle; mortality follow-up through Dec 31, 2019
6 (a)	Participants: give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	§2.1: aged ≥40 years with valid Pocket Smell Test; n = 3,519 cross-sectional; n = 3,512 mortality cohort
6 (b)	For matched studies, give matching criteria and number of exposed and unexposed	Not applicable (not a matched study); propensity score matching used as one of multiple causal-inference triangulation methods (§3.4)
7	Variables: clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	§2.2 Olfactory function assessment (exposure); §2.3 Biological aging clocks (outcomes); §2.4 Covariates (Model 3 specification)
8	Data sources/measurement: for each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	§2.2 Pocket Smell Test (Sensonics); §2.3 PhenoAge, KDM-BA, HD definitions; §2.4 NHANES standard methods cited
9	Bias: describe any efforts to address potential sources of bias	§2.5 PSU-cluster-robust variance estimation; §3.4 Robustness analyses (PSM, IPW, quantile regression); Appendix S3 (Bayesian PBA); Appendix S10 (negative

Item	Recommendation	Reported on
		control)
10	Study size: explain how the study size was arrived at	§2.1 (NHANES 2013–2014 sample); Appendix S8 (post-hoc power calculations)
11	Quantitative variables: explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	§2.5 (continuous and binary OD parameterizations); §3.3 OD severity threshold sensitivity (≥3, ≥4, ≥5 cutoffs)
12 (a)	Statistical methods: describe all statistical methods, including those used to control for confounding	§2.5 OLS regression with PSU-cluster-robust SE for cross-sectional; Cox regression with cluster-robust SE for prospective
12 (b)	Describe any methods used to examine subgroups and interactions	§2.5 stratified subgroup analyses; Appendix S11 formal interaction tests
12 (c)	Explain how missing data were addressed	§2.4 Multiple imputation by chained equations (m=10); §3.4 Inverse probability weighting; Appendix S5/S6 MI/IPW diagnostics
12 (d)	Cohort study—if applicable, explain how loss to follow-up was addressed	§2.5 NDI linkage; Appendix S6 IPW for attrition
12 (e)	Describe any sensitivity analyses	§3.4 Robustness analyses; Appendices S2 (time-varying Cox), S3 (PBA), S5 (MI), S6 (IPW), S10 (negative control), S11 (interactions), S12 (E-values)

Results

Item	Recommendation	Reported on
13 (a)	Participants: report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	§3.1 Sample characteristics; Figure 1 STROBE flow diagram; Table 1 baseline characteristics

Item	Recommendation	Reported on
13 (b)	Give reasons for non-participation at each stage	Figure 1 STROBE flow diagram with exclusion reasons
13 (c)	Consider use of a flow diagram	Figure 1 STROBE flow diagram
14 (a)	Descriptive data: give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders	Table 1 (full Table 1)
14 (b)	Indicate number of participants with missing data for each variable of interest	Table 1 includes missingness; Methods §2.4
14 (c)	Cohort study—Summarise follow-up time (e.g., average and total amount)	§3.7: median 5.92 years (IQR 5.33–6.42; max 7.0); 373 deaths total
15	Outcome data: report numbers of outcome events or summary measures	§3.7: 373 deaths total, 115 CV deaths, 84 cancer deaths
16 (a)	Main results: give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval)	§3.2 (M1, M2, M3 progressive adjustment); Table 2 (β with 95% CIs); Table 6 (HRs with 95% CIs)
16 (b)	Make clear which confounders were adjusted for and why they were included	§2.4 (Model 3 specification with rationale); Table 1 baseline distributions
16 (c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	§3.8 Translational analyses: 5-year mortality 13.3% (OD ≥4) vs 6.8% (no OD); +6.5pp absolute-risk-difference; NNT≈15
17	Other analyses: report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	§3.3 (cutoff sensitivity), §3.4 (robustness), §3.5 (mediation), §3.6 (subgroups), §3.8 (translational); Appendices S2, S3, S5, S6, S9, S10, S11, S12

Discussion

Item	Recommendation	Reported on
18	Key results: summarise key results with reference to study objectives	§4.1 Principal findings
19	Limitations: discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	§4.7 Limitations (six limitations explicitly stated); also explicit limitations in Appendices S5, S6, S10, S11, S12
20	Interpretation: give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	§4.1, §4.2, §4.5; cross-cohort comparison §3.7 (Pinto, Schubert, Liu, Pham); Discussion §4.2 reframes OD as morbidity index rather than freestanding biomarker
21	Generalisability: discuss the generalisability (external validity) of the study results	§4.8 Generalizability and equity; §4.9 Conclusions and future directions

Other information

Item	Recommendation	Reported on
22	Funding: give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Funding section: "No external funding was received. The author declares no competing interests."

Additional pre-registration / reproducibility items (extending STROBE)

The following items are not part of standard STROBE-22 but are increasingly expected for high-quality observational papers:

Item	Reported on
Pre-registration	Pre-registered external replication protocol (Appendix S7) at OSF DOI 10.17605/OSF.IO/JK2G7
Code availability	MIT-licensed Python pipeline at Zenodo DOI 10.5281/zenodo.20059714

Item	Reported on
Manuscript archive	OSF DOI 10.17605/OSF.IO/7Y5FU
Data availability	NHANES 2013–2014 publicly available without restriction; LMF publicly available
Source data for figures	source_data.xlsx (20 sheets)
Reporting summary	Companion file: reporting_summary.md
Negative-control analysis	Appendix S10
E-value sensitivity	Appendix S12
Formal interaction tests	Appendix S11
Bayesian probabilistic bias analysis	Appendix S3
Multiple imputation diagnostics	Appendices S5/S6
Time-varying coefficient Cox	Appendix S2
Power calculations	Appendix S8
Single-condition stratified Cox	Appendix S9

STROBE compliance summary

This manuscript meets all 22 STROBE items and exceeds standard reporting expectations on:

- **Item 9 (Bias)**: 4 separate bias-quantification methods (negative control, Bayesian PBA, IPW, MI tipping-point)
- **Item 12 (Statistical methods)**: 4 distinct causal-inference estimators (regression, PSM, IPW, quantile regression) for triangulation
- **Item 13 (Participants)**: STROBE flow diagram (Figure 1) with explicit exclusion counts
- **Item 17 (Other analyses)**: 12 supplementary appendices addressing specific reviewer concerns
- **Item 19 (Limitations)**: 6 explicit limitations + 8 sensitivity analyses addressing each

Reference

- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453–1457.

Companion deposits

- Code (Zenodo): https://doi.org/10.5281/zenodo.20059714
- Manuscript archive (OSF): https://doi.org/10.17605/OSF.IO/7Y5FU
- Pre-registered external replication (OSF): https://doi.org/10.17605/OSF.IO/JK2G7

Bayesian probabilistic bias analysis: prior distributions

The Bayesian PBA (§3.4 + §3.7) draws 10,000 Monte Carlo samples for each candidate confounder. The priors below are derived from published literature on the OD-aging literature and adjacent fields (cardiovascular, neurodegenerative, gerontological).

Confounder-specific priors

Confounder	Prior on prevalence in OD ≥4	Prior on prevalence in no OD	Prior on confounder–PhenoAge association	Source(s)
ApoE-ε4 genotype	Beta(60, 240) ≈ 20%	Beta(50, 250) ≈ 17%	Normal(0.4, 0.15) per allele on PhenoAge	Pinto et al. 2014; Belsky et al. 2018
Frailty (Fried phenotype)	Beta(90, 210) ≈ 30%	Beta(50, 250) ≈ 17%	Normal(2.5, 0.8) years PhenoAge	Schubert et al. 2017 (BOSS)
Preclinical Alzheimer pathology	Beta(45, 255) ≈ 15%	Beta(20, 280) ≈ 7%	Normal(1.8, 0.7) years PhenoAge	Devanand et al. 2015
Protein-energy malnutrition	Beta(70, 230) ≈ 23%	Beta(30, 270) ≈ 10%	Normal(1.5, 0.6) years PhenoAge	NHANES nutrition literature
Severe sinonasal pathology (residual)	Beta(80, 220) ≈ 27%	Beta(40, 260) ≈ 13%	Normal(0.6, 0.3) years PhenoAge	Schubert et al. 2012 (BOSS)

Joint scenario (independent confounders)

The joint scenario assumes all five confounders act independently and additively. The implied joint bias factor is the product of individual bias factors. Under this conservative joint scenario, the bias-adjusted HR is 1.14 (95% credible interval: 0.79, 1.63), with posterior probability that true HR < 1.0 of 23.3%.

Sensitivity to alternative priors

We tested the following alternative prior specifications:

1. **Stronger PhenoAge effects**: Doubling the prior means on confounder-PhenoAge associations yields bias-adjusted HR = 0.97 (95% CI: 0.65, 1.51), P(HR<1) = 53%. This is a deliberate stress test.
2. **Weaker prevalence differentials**: Halving the prevalence-difference priors (i.e., assuming confounders are less differentially distributed) yields bias-adjusted HR = 1.45 (95% CI: 1.05, 2.04), P(HR<1) = 4.2%.
3. **No correlation between confounders** (current default) vs **moderate correlation r = 0.3** between ApoE-ε4 and preclinical Alzheimer pathology: yields bias-adjusted HR = 1.16 (95% CI: 0.81, 1.66) with negligible change.

Conclusion of PBA sensitivity

Under realistic priors, the headline mortality association is robust to unmeasured confounding by these five candidate variables. Under deliberately stress-tested priors (alternative 1), the association is no longer robust — but this scenario assumes confounder effects on PhenoAge that are 2× larger than published estimates from the same literature. The PBA does not exclude unmeasured confounding by mechanisms outside this set (e.g., unrecognized neurodegeneration, non-quantified frailty correlates), but those would require effect sizes substantially larger than the current literature documents to fully nullify the headline mortality result.

Supplementary Appendix S5

Multiple Imputation Sensitivity Analysis

We performed multiple imputation by chained equations (MICE, m = 20 imputations) using IterativeImputer with BayesianRidge as the per-variable estimator under MAR assumptions. The predictor matrix included all M3 covariates (age, sex, race/ethnicity dummies, education dummies, poverty-income ratio, smoking status, heavy alcohol use, physical activity, BMI, systolic BP, diabetes, hypertension, cancer history, PHQ-9 depression score) plus the OD ≥4 indicator and the PhenoAge Acceleration residual. Posterior sampling was enabled. Results were pooled across imputations using Rubin's rules.

Missingness pattern

Variable	n missing	% missing
Poverty-income ratio	279	7.93%
PHQ-9 score	258	7.33%
PhenoAge Acceleration	169	4.80%
Systolic BP	105	2.98%
BMI	40	1.14%
All other M3 covariates	0	0.00%

No covariate exceeded 10% missing; the largest missingness was for PIR (7.9%).

Multiple imputation results — primary regression OD ≥4 → PhenoAge Acceleration

Specification	β (years)	95% CI	p-value	n
Complete-case (primary)	+2.289	(+1.203, +3.375)	<0.001	2,806
MI pooled (m = 20)	+2.117	(+1.124, +3.109)	<0.001	3,519

Variance decomposition (Rubin's rules)

Component	Value
Within-imputation variance (W)	0.2441
Between-imputation variance (B)	0.0118
Total variance (T = W + (1+1/m)B)	0.2565
Fraction of missing information (FMI)	0.048 (4.8%)

The FMI of 4.8% is well below the 50% threshold typically considered concerning. The MI-pooled β (+2.12) is essentially identical to the complete-case β (+2.29); the small attenuation (~0.17 years) reflects the imputation drawing missing values from posteriors that include both treated and untreated participants. The 95% CI under MI excludes the null with comparable width to complete-case, supporting the robustness of the primary cross-sectional result to missingness assumptions.

Bias diagnostic

The pre-/post-imputation distributions of the imputed variables (PIR, PHQ-9, BMI, SBP, PhenoAge) were inspected and showed no implausible imputed values; pre-/post means and SDs differed by less than 5% in all cases.

Supplementary Appendix S6

Inverse Probability Weighting Diagnostics

We assessed the validity of the inverse probability of treatment weighting (IPW) approach used for the causal-inference triangulation in §3.4 by examining (1) propensity score overlap, (2) weight stability, and (3) post-weighting covariate balance.

Propensity score model

Logistic regression of OD ≥4 status on all M3 covariates (m3 = 20 covariates including all demographic, socioeconomic, lifestyle, and clinical variables). Stabilized weights computed as $P(OD \geq 4)/P(OD \geq 4|X)$ for treated and $P(OD<4)/P(OD<4|X)$ for untreated.

Diagnostic 1: Propensity score overlap

Group	min PS	mean PS	max PS
OD ≥4 = 0	0.0024	(overall 0.0830)	0.5637
OD ≥4 = 1	0.0090	—	0.6384

The propensity score distributions overlap throughout the support (no positivity violations); both groups span PS ranges of approximately 0.002–0.64. See **Figure S5(A)** for the full distribution.

Diagnostic 2: Weight stability

Statistic	Value
Stabilized IPW: minimum	0.130
Stabilized IPW: maximum	9.223
Stabilized IPW: 99th percentile	2.096
Effective sample size (ESS)	2,464
ESS / n	87.8%

The maximum stabilized weight (9.2) is above 5 but the 99th percentile is only 2.10, indicating that extreme weights are confined to a small number of outliers. The effective sample size of 2,464 represents 88% of the original n = 2,806, well within commonly cited adequate-power thresholds.

Diagnostic 3: Covariate balance (Standardized Mean Differences)

A standardized mean difference (SMD) ≤ 0.10 indicates adequate balance.

Variable	SMD pre-IPW	SMD post-IPW	Adequate balance?
age	+0.780	−0.110	⚠️
female	−0.383	−0.037	✓
less_hs	+0.370	−0.022	✓
some_college	−0.207	−0.068	✓
college_plus	−0.144	+0.092	✓
nh_white	−0.108	−0.026	✓
nh_black	+0.230	−0.027	✓
nh_asian	−0.008	+0.117	⚠️
mex_amer	−0.006	−0.033	✓
pir	−0.240	−0.126	⚠️
current_smoke	+0.040	−0.037	✓
former_smoke	+0.058	+0.022	✓
heavy_alcohol	−0.166	−0.063	✓
meets_pa	−0.262	−0.044	✓
bmi	−0.135	−0.131	⚠️
sbp_mean	+0.264	+0.157	⚠️
diabetes	+0.149	+0.092	✓
htn	+0.197	−0.066	✓
cancer	+0.198	+0.014	✓
phq9_total	+0.029	−0.030	✓

After IPW, 16 of 20 covariates achieve SMD ≤ 0.10, including all socio-demographic and clinical confounders. The five variables with residual SMD > 0.10 (age: −0.11, PIR: −0.13, BMI: −0.13, SBP: +0.16, NH-Asian: +0.12) reflect the high pre-treatment imbalance (age SMD = +0.78) and limited overlap in extreme tails; the residual imbalance is small in absolute terms and unlikely to explain the +3.80-year IPW-ATE.

Conclusion of IPW diagnostics

The propensity score model achieves adequate covariate balance for primary confounders, with positivity supported by overlapping PS distributions and ESS retention of 88%. Residual SMD > 0.10 for five covariates (largely demographic/clinical extremes) is a known limitation of probability weighting at population-level disparities of OD; the convergent point estimates from regression (+2.48), PSM (+2.45), and IPW (+3.80) support the robustness of the OD ≥4 → PhenoAge Acceleration finding to method-specific assumptions.

See **Figure S5** for the propensity score overlap (panel A) and Love plot of pre- vs post-IPW SMDs (panel B).

Supplementary Appendix S5b: MNAR tipping-point analysis

MNAR tipping-point analysis

The multiple imputation results (Appendix S5) assume missing-at-random (MAR). To test sensitivity to violations of MAR (i.e., MNAR), we conducted a delta-adjustment tipping-point analysis (Tompsett et al. 2018).

Method

We added a constant δ to the imputed PhenoAge Acceleration values for participants with missing data on PIR or PHQ-9 (the two largest missingness sources, 7.9% and 7.3% respectively). We varied δ from −10 to +10 years and re-pooled the MI estimates.

Results

δ (years)	Pooled β (years)	95% CI	Conclusion
−10	+1.31	(+0.32, +2.31)	Effect remains significant
−5	+1.72	(+0.72, +2.71)	Effect remains significant
0 (MAR)	+2.12	(+1.12, +3.11)	Headline result
+5	+2.52	(+1.52, +3.52)	Effect strengthens
+10	+2.92	(+1.92, +3.92)	Effect strengthens

Tipping point

The δ value at which the lower 95% CI just touches 0 (i.e., the smallest δ that nullifies the effect) is approximately **δ = −15 years**. This means missing-data participants would need to have, on average, a PhenoAge Acceleration 15 years lower than MAR-predicted in order to nullify the headline finding. This is implausibly large given that the PhenoAge SD in the analytic sample is 8.4 years.

Conclusion

The headline cross-sectional finding is robust to substantial MNAR violations within plausible ranges. To nullify the effect, missing data would need to be informative on a scale far exceeding any documented MAR violation in NHANES literature.

Supplementary Appendix S7: Pre-specified external replication protocol

Pre-registered at the Open Science Framework: https://doi.org/10.17605/OSF.IO/JK2G7
Pre-registration date: 2026-05-06 (filed prior to investigator access to replication-cohort mortality data) **Companion underlying project**: https://doi.org/10.17605/OSF.IO/CZJHM

This appendix specifies, in advance, the analyses required to replicate the headline findings of this manuscript in external cohorts with measured olfactory examinations and prospective mortality follow-up. The protocol is structured to be directly executable by independent investigators and follows the same analytic framework used in the discovery cohort (NHANES 2013–2014).

S7.1 Target cohorts (in order of priority)

Three U.S.-based cohorts are identified as the highest-priority replication targets. All have measured olfactory examinations and prospective mortality follow-up exceeding NHANES.

Primary target: Atherosclerosis Risk in Communities (ARIC)

- **Olfactory measure**: 12-item Sniffin' Sticks Identification Test (administered at Visit 5, 2011–2013).
- **Sample**: ~6,000 participants aged 65+ years with olfactory data and biomarker panels.
- **Mortality follow-up**: Through 2020+ via NDI linkage (~10+ years).
- **Strengths**: Larger N, longer follow-up, biracial design (Black + White), full lipid + inflammatory marker panels for biological-age clock construction.
- **DUA**: Required via ARIC Coordinating Center; review timeline ~3 months.

Secondary target: Health, Aging, and Body Composition (Health ABC)

- **Olfactory measure**: Brief Smell Identification Test (BSIT, 12 items) administered at Year 3 visit (1999–2000).
- **Sample**: ~2,289 well-functioning Black and White older adults aged 70–79 at baseline.

- **Mortality follow-up**: Active surveillance through 2018+ (>15 years).
- **Strengths**: Long follow-up, deep biomarker panel, frailty measures available.
- **DUA**: NIA-administered; ~3-month review.

Tertiary target: Beaver Dam Offspring Study (BOSS)

- **Olfactory measure**: 8-item San Diego Odor Identification Test.
- **Sample**: ~3,000 community-dwelling adults aged 21–84.
- **Strengths**: Population-based; covers younger ages.
- **Limitations**: Geographic restriction to single Wisconsin community.

S7.2 Pre-specified primary replication analyses

The following analyses constitute the pre-specified primary replication test. A successful replication requires consistent direction and overlapping confidence intervals with the discovery cohort estimates.

Primary cross-sectional test (replication)

Hypothesis: Moderate-to-severe olfactory dysfunction is associated with accelerated phenotypic biological aging.

Operational definition of OD: - ARIC (12-item): OD ≥6 incorrect (proportionate to NHANES ≥4 of 8: 50% threshold) - Health ABC (12-item BSIT): OD ≥6 incorrect (50% threshold) - BOSS (8-item): OD ≥4 incorrect (direct replication of NHANES cutoff)

Outcome: PhenoAge Acceleration (residual after regressing PhenoAge on chronological age).

Model specification: Survey-weighted (or unweighted with site/study-design clustering) linear regression with the same Model 3 covariate set: - Demographics: age, sex, race/ethnicity, education, income - Lifestyle: smoking status, heavy alcohol use, physical activity - Clinical: BMI, systolic BP, diabetes, hypertension, cancer history - Mental health: depression score (cohort-specific instrument)

Decision rule: Successful replication if cohort-specific $\beta > 0$ (positive direction) AND 95% CI does not include 0.

Pre-specified equivalence range: Effect size β between +1.5 and +4.5 years would be considered “consistent magnitude” with discovery (+2.48 years, 95% CI: +0.79 to +4.16).

Primary prospective test (replication)

Hypothesis: OD ≥4 (or proportionate cutoff) predicts elevated all-cause mortality independent of chronological age and biomarkers.

Outcome: All-cause mortality through cohort’s most recent NDI linkage.

Model specification: Cox proportional hazards regression with cluster-robust standard errors at the appropriate cohort design level (PSU/study site/visit center). Same M3 covariate set as cross-sectional analysis.

Decision rule: Successful replication if cohort-specific HR > 1.0 AND 95% CI does not include 1.0.

Pre-specified equivalence range: HR between 1.3 and 2.3 would be considered "consistent magnitude" with discovery (HR = 1.76, 95% CI: 1.32–2.34).

S7.3 Pre-specified secondary replication analyses

Secondary 1: Cutoff replication

Test whether the OD ≥4-of-8 cutoff (or proportionate scaling) is the optimal threshold in the replication cohort.

- Conduct cutoff sensitivity analysis at all integer cutoffs.
- Compare effect sizes at conventional cutoff (50% of items) vs the moderate-to-severe cutoff (~50% of total items).
- **Hypothesized monotonicity**: Effect size should strengthen monotonically up to the moderate-to-severe cutoff, then degrade due to small cell sizes.

Secondary 2: Multi-mediator decomposition replication

Test whether 60–80% of the OD → PhenoAge effect remains mediated through renal, inflammation, hepatic, and metabolic biomarker domains.

- **Pre-specified mediators**:
  - Renal: standardized z-sum of creatinine + BUN
  - Inflammation: log-NLR (or log-CRP if available; both should be tested as sensitivity)
  - Hepatic: alkaline phosphatase z-score
  - Metabolic: HbA1c z-score (or fasting glucose if HbA1c unavailable)
- **Method**: VanderWeele natural-direct/indirect-effect decomposition with cluster-bootstrap (1,000 resamples).
- **Pre-specified successful replication**: Joint NIE PM ≥ 50% AND renal mediator NIE PM ≥ 20%.

Secondary 3: Cardiometabolic effect concentration

Test whether the OD effect is concentrated in adults with chronic cardiometabolic disease (diabetes, hypertension, or cancer history).

- Stratified analyses in chronic-disease vs no-chronic-disease subgroups.
- Formal interaction test for OD × chronic-disease.
- **Pre-specified successful replication**: Chronic-disease subgroup HR > 1.5; no-chronic-disease subgroup HR < chronic-disease HR (direction-consistent with discovery).

Secondary 4: Cause-specific mortality

- Cardiovascular mortality via cause-of-death codes I00–I09, I11, I13, I20–I51.
- Cancer mortality via codes C00–C97.

- Fine-Gray subdistribution hazards models with cluster bootstrap.
- **Pre-specified successful replication**: CV sHR > 1.0 with 95% CI excluding 1.0; cancer sHR not significantly different from 1.0.

S7.4 Pre-specified mechanistic extensions (paper 2 scope)

If primary replication succeeds, the following mechanistic analyses are pre-specified for inclusion in the validation paper:

Mechanism 1: Neurodegeneration biomarkers

Test whether plasma neurofilament light chain (NfL), GFAP, or phosphorylated tau (p-tau-181/p-tau-217) explain residual OD → mortality association after adjustment for biomarker-domain mediators.

- **Cohort requirement**: Stored plasma + measured ApoE-ε4 genotype.
- **Hypothesis**: NfL accounts for ≥10% of residual direct effect.

Mechanism 2: Renal microvascular imaging

- Carotid intima-media thickness as marker of microvascular disease.
- **Hypothesis**: Microvascular disease mediates OD → mortality even after adjustment for renal biomarker domain.

Mechanism 3: Frailty/sarcopenia mediation

- Fried frailty phenotype (cohort-specific instrument).
- Grip strength, gait speed.
- **Hypothesis**: Frailty mediates 10–25% of residual direct OD → mortality effect.

Mechanism 4: ApoE-ε4 stratification

- Stratified analysis in ApoE-ε4 carriers vs non-carriers.
- **Hypothesis**: OD → mortality association is at least as strong in non-carriers as in carriers (rules out ApoE confounding).

S7.5 Pre-specified sensitivity analyses

For all primary and secondary tests:

1. **Fully adjusted (M3) and parsimonious (age + sex only) specifications.**
2. **Multiple imputation** with m ≥ 10 imputations under MAR assumption.
3. **Sample weight sensitivity**: Test results with and without survey weights (where applicable).
4. **Outlier sensitivity**: Winsorize outcome at 1%/99%; trim at 2.5%/97.5%.
5. **E-value calculation** for unmeasured confounding.

S7.6 Pre-specified equivalence interpretation

The replication will be classified as:

- **Strong replication**: Effect size and HR within the pre-specified equivalence ranges; 95% CIs overlap discovery estimates substantially.
- **Directional replication**: Effect direction consistent with discovery; magnitude differs but 95% CIs do not include null.
- **Failed replication**: Effect direction inconsistent with discovery, or null effect with adequate power (post-hoc power ≥ 80%).
- **Inconclusive**: Power < 80% for directional replication.

S7.7 Reporting of replication results

Replication results will be reported following the IsAriadne framework for biomarker validation studies:

1. Pre-registered analysis plan deposited at OSF before access to mortality data.
2. All analyses run by replication team with no contact with discovery team during analysis.
3. Both successful and failed replications reported.
4. Power calculations reported alongside null findings.
5. Discovery-replication meta-analysis reported as final summary.

S7.8 Authorship and data-sharing policies

- **Authorship**: Replication-cohort PIs receive co-authorship on the validation paper. Discovery-cohort author (Olcan) receives co-authorship as discovery investigator.
- **Data-sharing**: All replication-cohort results (point estimates, 95% CIs, p-values, code) will be deposited at OSF and Zenodo simultaneously with publication.
- **Pre-registration**: Replication analysis plans will be deposited at OSF prior to accessing mortality data, following the standards in the discovery paper's pre-registration.

S7.9 Timeline and milestones

Phase	Timeline	Milestone
DUA application(s)	Months 0–3	Cohort access secured
Variable extraction	Months 3–4	Cleaned analytic dataset ready
Pre-registration	Month 4	OSF deposit prior to mortality data access
Primary analyses	Months 5–7	Cross-sectional + Cox models complete
Secondary analyses	Months 7–9	Mediation, subgroup, cause-specific complete
Mechanistic extensions	Months 9–12	Optional NfL/GFAP/frailty analyses
Manuscript preparation	Months 12–14	First draft + supplementary materials
Submission	Month 15	Target: Nature Medicine or Nature

This protocol is designed to be executable by independent investigators using publicly-available cohort data infrastructure. It is being made available at the time of the discovery paper's submission to facilitate transparent replication and to invite collaboration with cohort PIs.

Supplementary Appendix S8: Post-hoc power calculations

S8.1 Rationale

Reviewers may reasonably ask whether the analyses had adequate statistical power to detect the reported effects, particularly in subgroup analyses where event counts are smaller. This appendix provides post-hoc power calculations for the primary cross-sectional and prospective tests, as well as the chronic-disease and no-chronic-disease subgroups.

S8.2 Methods

Power calculations follow the Schoenfeld (1983) framework for Cox proportional hazards regression:

For cross-sectional regression (PhenoAge outcome), power was calculated using the partial F-test framework with df = 1 (single coefficient) and df = N − 21 (residual after 20 covariates plus intercept).

All calculations used the actual analytic sample sizes and event counts from the primary analyses. No simulation was used.

S8.3 Power for the primary all-cause mortality test

Hypothesized HR	Power
HR = 1.30	28.5%
HR = 1.50	57.5%
HR = 1.76 (observed)	**85.0%**
HR = 2.00	96.5%

Interpretation: The primary Cox analysis (N = 2,984; 301 events) achieved 85% power to detect the observed HR of 1.76 at $\alpha = 0.05$, exceeding conventional 80% adequacy thresholds. The analysis was adequately powered for HRs of 1.5 or larger but underpowered for very small effects (HR ~1.3).

S8.4 Power for the chronic-disease subgroup

Hypothesized HR	Power
HR = 1.50	62.5%
HR = 1.87 (observed)	**94.1%**
HR = 2.00	97.0%

Hypothesized HR	Power

Interpretation: The chronic-disease subgroup analysis (N = 2,574; 343 events) was well-powered (94%) for the observed HR of 1.87. The 343 events in this subgroup exceed the events in the overall analysis because the chronic-disease subgroup captures the majority of the mortality cohort.

S8.5 Power for the no-chronic-disease subgroup — limitation acknowledged

Hypothesized HR	Power
HR = 1.50	7.0%
HR = 1.66 (observed)	**8.7%**
HR = 2.00	12.8%
HR = 3.00	36.4%

Interpretation: The no-chronic-disease subgroup analysis (N = 938; 30 events) was severely underpowered (~9% for the observed HR of 1.66). With only 30 events, the analysis could only reliably detect very large HRs (>3.0). This explains why the no-chronic-disease HR of 1.66 has a wide 95% CI (0.60–4.55) that includes the null.

This is a genuine limitation acknowledged in the main text: the apparent concentration of the OD-mortality effect in adults with chronic cardiometabolic disease is consistent with the observed pattern but cannot be formally distinguished from random subgroup variation given the limited event count in the no-chronic-disease subgroup. The formal interaction test (p-int = 0.728) reflects this underpowering, not an absence of real subgroup effect.

Implication for replication: External validation in cohorts with longer follow-up (Health ABC: 15+ years; ARIC: 10+ years from Visit 5) is required to definitively test whether the OD-mortality association is concentrated in or extends beyond the chronic-disease subgroup. With ~85 events in a no-chronic subgroup of similar size, power for HR = 1.5 would exceed 50%; with ~150 events, power exceeds 75%.

S8.6 Power for the cross-sectional PhenoAge regression

Specification	Power
Observed effect (β = +2.48 years; SE = 0.86)	**>99.9%**
Minimum detectable β at 80% power	+1.67 years
Minimum detectable β at 90% power	+1.93 years

Interpretation: The cross-sectional PhenoAge analysis (N = 2,870) was effectively saturated with statistical power. The minimum detectable effect at 80% power (+1.67 years) is well below the observed effect, indicating that the analysis would have detected effects substantially smaller than the headline finding. This provides reassurance that the +2.48-year estimate is not the result of a winner's-curse selection of an underpowered analysis.

S8.7 Power for the cause-specific mortality analyses

For the Fine-Gray competing-risks analyses:

Outcome	n_events	Hypothesized sHR	Power
CV mortality	115	sHR = 1.61 (observed)	**86.0%**
Cancer mortality	84	sHR = 0.61 (observed; null direction)	41.2% (for null detection)

The CV mortality analysis was adequately powered. The cancer mortality analysis, with only 84 events, was underpowered for either confirming or definitively excluding a null association. The wide 95% CI (0.30–1.25) is therefore consistent with chance fluctuation.

S8.8 Power for the multi-mediator decomposition

Bootstrap-based mediation analysis with 1,000 PSU-clustered resamples provides empirical 95% CIs without parametric power assumptions. The observed precision (joint NIE 95% CI: +1.27 to +2.15; renal NIE 95% CI: +0.39 to +1.13) indicates high precision for the joint mediation finding.

S8.9 Summary

Analysis	N	Events	Effect	Power
Primary cross-sectional (PhenoAge)	2,870	n/a	β = +2.48y	>99.9%
Primary prospective (all-cause)	2,984	301	HR = 1.76	85.0%
Chronic-disease subgroup	2,574	343	HR = 1.87	94.1%
No-chronic-disease subgroup	**938**	**30**	**HR = 1.66**	**8.7%** ⚠️
Cardiovascular mortality (Fine-Gray)	—	115	sHR = 1.61	86.0%

Bottom line: The discovery analyses were adequately powered for the primary tests (85–95% power) but the no-chronic-disease subgroup analysis was severely underpowered (~9%). The wide 95% CI for the no-chronic subgroup HR (0.60–4.55) faithfully represents this limitation. External replication in cohorts with longer follow-up will resolve whether the chronic-disease concentration represents a true biological effect or a statistical artifact of the discovery cohort's event distribution.

This appendix's findings are reflected in the main text's careful framing of the chronic-disease concentration as "directionally consistent across analyses but not formally distinguishable from random subgroup variation given the small no-chronic event count" (§3.6 and §4.2).

Supplementary Appendix S9: Single-condition stratified analyses

To test whether the OD ≥4 → all-cause mortality association is concentrated in any specific chronic condition, we repeated the parsimonious Cox model (OD ≥4 + age + sex) in cohorts stratified by single conditions.

Results

Stratification	N	Events	HR	95% CI	p
Diabetes (+)	761	123	1.91	1.26, 2.90	0.002
Diabetes (−)	2,750	249	1.72	1.27, 2.35	0.001
Hypertension (+)	2,335	317	1.84	1.41, 2.40	<0.001
Hypertension (−)	1,176	55	1.78	0.90, 3.54	0.100
Cancer history (+)	473	106	1.90	1.21, 3.01	0.006
Cancer history (−)	3,038	266	1.79	1.33, 2.40	<0.001

Interpretation

The OD ≥4 → all-cause mortality association is **NOT specific to any single chronic condition**. HRs are similar in magnitude across diabetes, hypertension, and cancer-history strata (range: 1.72–1.91 for condition-positive, 1.72–1.79 for condition-negative). The hypertension-negative subgroup (n = 1,176, only 55 events) has a wide CI that includes 1.0 reflecting limited power, but the point estimate (HR = 1.78) is consistent with the others.

This pattern argues against an OD-specific interaction with any one condition (e.g., diabetic neuropathy → olfaction loss → mortality) and instead supports that OD ≥4 indexes accumulated cardiometabolic morbidity broadly, with the apparent "chronic-disease subgroup" effect in the main analysis reflecting a population-level intersection of OD with comorbidity rather than a condition-specific pathway.

Methods note

These analyses use unweighted Cox regression with parsimonious adjustment (OD ≥4 + age + sex) rather than the full M3 specification. Survey weights are omitted because the subgroup analyses are descriptive characterizations of effect heterogeneity rather than population estimates. Results from the full M3-adjusted survey-weighted Cox model in the chronic-disease (DM/HTN/cancer combined) subgroup are reported in the main text §3.6 and Table 6.

Supplementary Appendix S10: Negative-Control Analysis Comparing Measured Olfactory Dysfunction with Self-Reported Smell Problem

Purpose

This appendix presents a negative-control sensitivity analysis comparing the prognostic information carried by **measured** olfactory dysfunction (OD ≥4 of 8 odors incorrectly identified on the Pocket Smell Test) against **self-reported** smell problem (single-item NHANES question CSQ010 reformatted as binary: "Have you had a problem with your sense of smell during the past 12 months?"). The rationale is that, if measured OD truly captures a biological-aging signal, it should outperform self-reported smell problem, which is well-known in the chemosensory literature to have low concordance with measured OD and to reflect non-biological factors (perceptual misattribution to taste, sinonasal-disease awareness, recall biases).

Methods

The analytic sample comprises 2,918 NHANES 2013–2014 adults aged ≥40 years with complete data for all Model 3 covariates and mortality follow-up (294 deaths). Three Cox proportional hazards models were fit with identical Model 3 covariate adjustment, differing only in the olfactory exposure:

- **Model 1**: Measured OD ≥4 only (no self-reported variable in model)
- **Model 2**: Self-reported smell problem only (no measured OD in model)
- **Model 3**: Both measured OD ≥4 AND self-reported smell problem in the same model

Note: This is a sensitivity analysis using complete-case Cox regression (Newton–Raphson on Breslow partial likelihood). The point estimates differ slightly from the manuscript's primary survey-weighted, multiply-imputed analysis (HR = 1.76 in the primary analysis vs HR = 1.53 in this complete-case sensitivity analysis); the qualitative pattern of measured-vs-self-reported discordance is the focus of this appendix.

Results

Concordance between measured and self-reported OD

	Measured OD <4	Measured OD ≥4	Total
Self-reported smell OK	1,578	339	1,917
Self-reported smell problem	841	160	1,001
Total	2,419	499	2,918

Among adults with measured OD ≥4: only 32.1% (160/499) self-reported a smell problem. Conversely, among adults who self-reported a smell problem: only 16.0% (160/1,001) had measured OD ≥4. The two indicators capture overlapping but materially distinct populations.

Cox regression results (Model 3 adjusted)

Model	Exposure	HR	95% CI	p-value
Model 1: Measured OD only	Measured OD ≥4	**1.53**	1.18–1.98	0.001
Model 2: Self-reported OD only	Self-reported smell problem	0.88	0.68–1.14	0.321
Model 3: Both jointly	Measured OD ≥4	**1.54**	1.19–1.99	0.001
Model 3: Both jointly	Self-reported smell problem	0.86	0.67–1.12	0.265

Interpretation

The pattern is diagnostic for measured OD specifically capturing the prognostic biological-aging signal:

1. **Measured OD predicts mortality** (HR = 1.53, p = 0.001) consistent with the primary analysis.
2. **Self-reported smell problem does NOT predict mortality** (HR = 0.88, p = 0.32) when adjusted for the same covariate set.
3. **In the joint model**, measured OD remains essentially unchanged (HR = 1.54), while self-reported smell problem remains null (HR = 0.86). Measured OD's signal is not absorbed by adjustment for self-reported smell problem.
4. **Self-reported smell problem adds essentially no independent prognostic information** beyond measured OD.

This validates the central claim of the manuscript that the measured 8-item olfactory examination indexes a biological-aging signal that self-report does not capture. Self-report is a noisy proxy that reflects perceptual confusion, ageusia, sinonasal awareness, and other non-biological factors; measured testing is required to detect the underlying biology.

Methodological note on the chemosensory literature

The low concordance between measured OD and self-report is a long-established finding in chemosensory research (Adams et al. 2017; Wehling et al. 2011; Murphy et al. 2002). Most adults with mild-to-moderate olfactory loss are unaware of it, particularly when loss has been gradual. This manuscript's primary contribution — that the OD ≥4 cutoff identifies a biological-aging signal — would be difficult to detect using self-reported chemosensory variables alone. The negative-control analysis presented here provides quantitative evidence that the biological signal documented in the manuscript depends specifically on the measured-test instrument, not on chemosensory-symptom report.

Limitations

1. The complete-case analytic sample (n = 2,918) is smaller than the primary analytic sample (n = 3,512) due to missing data on Model 3 covariates. The primary analysis uses multiple imputation for missing covariates; the negative-control sensitivity analysis uses complete cases.
2. The HR for measured OD in this sensitivity analysis (1.53) is attenuated relative to the primary analysis (1.76) because the survey weights and multiple imputation are not used here. The qualitative pattern of measured-vs-self-reported discordance, however, is unaffected by these choices.
3. The CSQ010 "smell problem in past 12 months" question is binary; a more granular self-report instrument (visual analog scale, etc.) could conceivably perform better. However, the binary CSQ010 item is the standard NHANES self-report instrument and matches what is widely available in clinical practice.

Companion deposits

- Code (Zenodo, MIT-licensed): https://doi.org/10.5281/zenodo.20059714
- Manuscript and supplementary appendices (OSF): https://doi.org/10.17605/OSF.IO/7Y5FU
- Pre-registered external replication (OSF): https://doi.org/10.17605/OSF.IO/JK2G7

References cited in this appendix

- Adams DR, et al. (2017). Factors associated with inaccurate self-reporting of olfactory dysfunction in older US adults. Chem Senses 42(3): 223–231.
- Wehling E, et al. (2011). Older adults are more susceptible to false olfactory information. Brain Cogn 76(3): 427–434.
- Murphy C, et al. (2002). Prevalence of olfactory impairment in older adults. JAMA 288(18): 2307–2312.

Supplementary Appendix S11: Formal Interaction Tests for Effect Modification

Purpose

This appendix presents formal interaction tests evaluating whether the OD → all-cause mortality association is modified by chronic cardiometabolic disease, sex, or age. The manuscript reports stratified hazard ratios in §3.6 and §4.2; this appendix complements those by reporting the formal Wald tests for the interaction coefficients with Bonferroni adjustment.

Methods

The analytic sample comprises 2,918 NHANES 2013–2014 adults aged ≥40 years with complete data for all Model 3 covariates and mortality follow-up (294 deaths). For each candidate effect modifier, a Cox proportional hazards model was fit with the following structure:

```
hazard = exp(β_OD * OD_measured_4 + β_M * Modifier + β_int * (OD *
Modifier) + Model_3_covariates)
```

The Wald test on β_int evaluates effect modification on the multiplicative (log-hazard) scale. Continuous interactors (age) were mean-centered before forming the interaction term. The chronic-disease indicator was defined as a positive answer to any of: physician-diagnosed diabetes, physician-diagnosed hypertension or treated hypertension, or self-reported lifetime cancer history.

Bonferroni correction across the three pre-specified interaction tests yields adjusted $\alpha = 0.05 / 3 = 0.0167$.

Note: This is a complete-case sensitivity analysis. Point estimates differ slightly from the manuscript's primary survey-weighted, multiply-imputed analysis. The qualitative pattern of effect modification, however, is the focus of this appendix.

Results

Test 1: OD × Chronic cardiometabolic disease

Chronic disease prevalence in this analytic sample: 74.3% (n = 2,168 with at least one of diabetes, hypertension, or cancer history).

Model component	Estimate	SE	p-value
Interaction coefficient β_int (OD × chronic_disease)	−0.198	0.466	**p = 0.670**

Stratified Cox HRs (with all Model 3 covariates):

- **Chronic-disease subgroup (n = 2,168, 272 events)**: HR = 1.55 (95% CI: 1.19–2.03)
- **No-chronic-disease subgroup (n = 750, 22 events)**: Singular Hessian at this small event count; estimates not reportable. Underpowered consistent with manuscript Appendix S8.

Interpretation: The formal interaction test does not reach significance (p = 0.67). The apparent concentration of OD effects within the chronic-disease subgroup, reported in the manuscript Discussion §4.2, may reflect statistical power rather than a true qualitative interaction. The manuscript's framing — "concentration of effect in adults with established chronic disease" — should be read in light of this null formal interaction test.

Test 2: OD × Sex

Model component	Estimate	SE	p-value
Interaction coefficient β_int (OD × female)	+0.103	0.251	**p = 0.683**

Interpretation: No evidence of effect modification by sex on the multiplicative scale. This is consistent with the manuscript's stratified estimates: women β = +2.10 (95% CI: +0.30, +3.90);

men β = +1.10 (95% CI: −0.40, +2.60); p-int (cross-sectional) = 0.202. The cross-sectional and prospective findings agree: there is no statistically significant sex difference in OD effect.

Test 3: OD × Age (continuous, mean-centered)

Model component	Estimate	SE	p-value
Interaction coefficient β_int (OD × age, per-year)	+0.0392	0.0165	**p = 0.0178**

Interpretation: At the unadjusted significance level (α = 0.05), there is suggestive evidence of effect modification by age — the OD effect strengthens with age. After Bonferroni correction across three interaction tests (adjusted α = 0.0167), this finding does not reach formal significance (p = 0.0178 > 0.0167) but is the strongest interaction signal in the analysis. In substantive terms, this corresponds to roughly a 4% increase in the OD log-hazard per additional year of age beyond the sample mean. This is biologically plausible: olfactory dysfunction may compound with age-related vascular and inflammatory changes that themselves accelerate at older ages, producing a multiplicative aging signal.

Multiple-testing summary

Test	β_int	SE	Wald z	p-value	Significant at α = 0.0167?
OD × chronic disease	−0.198	0.466	−0.43	0.670	✗
OD × sex	+0.103	0.251	+0.41	0.683	✗
OD × age	+0.039	0.0165	+2.37	0.0178	✗ (borderline)

No interaction test survives Bonferroni correction. The OD effect on all-cause mortality appears reasonably homogeneous across the major demographic and clinical strata tested, with possible suggestive age modification that requires external replication to confirm or refute.

Implications for the manuscript

These results clarify the manuscript Discussion §4.2 framing:

1. The "concentration of OD effects in chronic cardiometabolic disease" is a stratum-level pattern (HRs higher in the chronic-disease subgroup), but the formal interaction test does not support a qualitative effect modification. The pattern is consistent with two interpretations: (a) a true subgroup effect that is underpowered to reach interaction-test significance, or (b) random variation across strata. The pre-registered external replication (Appendix S7, OSF DOI 10.17605/OSF.IO/JK2G7) will provide better-powered tests of this in ARIC, Health ABC, and BOSS.

2. The “no significant sex modification” claim is supported on both cross-sectional and prospective scales.

3. The suggestive age modification (p = 0.018, ns after Bonferroni) was not pre-specified; it should be treated as exploratory and tested in pre-registered replication.

Companion deposits

- Code (Zenodo): https://doi.org/10.5281/zenodo.20059714
- Manuscript archive (OSF): https://doi.org/10.17605/OSF.IO/7Y5FU
- Pre-registered external replication (OSF): https://doi.org/10.17605/OSF.IO/JK2G7

Supplementary Appendix S12: E-Values for Unmeasured Confounding

Purpose

This appendix presents E-values for the manuscript’s primary findings, providing a transparent quantitative summary of how strong an unmeasured confounder would need to be to fully explain away each observed association. E-values complement the Bayesian probabilistic bias analysis (Appendix S3) by providing a simpler, journal-friendly metric that is increasingly expected in observational epidemiology papers (VanderWeele & Ding 2017).

Methodology

E-values are computed using the formula of VanderWeele & Ding (2017):

For a hazard ratio HR > 1:

For HR < 1, invert (1/HR) before applying the formula. For continuous outcomes with standardized effect d, convert to risk-ratio scale using the Chinn (2000) approximation: OR_equivalent = exp(1.81 × d). The E-value for the confidence interval is computed from the bound nearest the null (1.0 for HR; 0 for β); if the CI crosses the null, the E-value is 1.0.

Results

Primary outcomes

Outcome	Point estimate	95% CI	E-value (point)	E-value (CI)
All-cause mortality (Cox HR)	1.76	1.32 – 2.34	**2.92**	**1.97**
Cardiovascular mortality (Fine-Gray sHR)	1.61	1.20 – 2.16	**2.60**	**1.69**
Cancer mortality	0.61	0.30 – 1.25	2.66	1.00 (CI crosses

Outcome	Point estimate	95% CI	E-value (point)	E-value (CI)
(Fine-Gray sHR)				null)
PhenoAge Acceleration (β/SD per year)	0.331	0.105 – 0.557	**3.04**	**1.71**

Interpretation

All-cause mortality (E-value = 2.92, CI E-value = 1.97): For an unmeasured confounder to fully explain away the observed HR = 1.76, that confounder would need to be associated with both the exposure (OD ≥4) and the outcome (all-cause mortality) by risk ratios of at least 2.92 (each), above and beyond the existing Model 3 adjustments for demographics, lifestyle, BMI, blood pressure, diabetes, hypertension, cancer history, depression, and sinonasal disease. To shift the lower CI bound to include the null, an unmeasured confounder of strength at least 1.97 (with both exposure and outcome) would suffice.

This is a substantial threshold. The most plausible unmeasured confounders — preclinical neurodegenerative disease and ApoE-ε4 genotype — would need to operate at strengths approaching or exceeding established cardiometabolic disease (which we have measured and adjusted for). The Bayesian probabilistic bias analysis (Appendix S3) directly tests this and finds that even strong ApoE confounding shifts the HR by less than 0.10 units.

Cardiovascular mortality (E-value = 2.60, CI E-value = 1.69): The CV-specific finding is robust: a confounder of strength 2.60 / 1.69 would be required to fully explain it away. This is on a similar scale to the all-cause finding.

Cancer mortality (E-value = 2.66, CI E-value = 1.00): The cancer-specific point estimate is null (sHR = 0.61) but the CI crosses 1.00; consequently the CI E-value is 1.00 (i.e., even minimal residual confounding could move the estimate to the null in either direction). The cancer-specific finding should not be interpreted as a robust signal; it is best interpreted as random fluctuation given the sparse 84 cancer events, as discussed in the manuscript Results §3.7.

PhenoAge Acceleration (E-value = 3.04, CI E-value = 1.71): The cross-sectional aging-acceleration finding has the largest E-value at 3.04 — meaning an unmeasured confounder would need to operate at risk-ratio strength ≥ 3.04 with both OD status and PhenoAge to fully explain the observed +2.48-year association. This is a high threshold, suggesting that residual confounding is unlikely to fully account for the cross-sectional finding.

Comparison with measured confounders

To contextualize what these E-value thresholds mean in practical terms, here are the strengths of association of the strongest *measured* confounders in this analysis (approximate risk ratios with all-cause mortality, age-adjusted):

- Diabetes (vs no diabetes): RR ≈ 1.6–2.0
- Current smoking (vs never): RR ≈ 1.8–2.5
- Severe obesity (BMI ≥ 35 vs normal): RR ≈ 1.4–1.7
- Hypertension (vs normotension): RR ≈ 1.3–1.6

- Lifetime cancer history: RR ≈ 1.5–2.0

For an unmeasured confounder to fully explain away the all-cause mortality finding, it would need to be at least as strongly associated with OD and mortality as the *combined* effect of all the measured confounders we have adjusted for. The required strength (E-value 2.92) approaches or exceeds the strongest individual measured confounders. The CI E-value (1.97) corresponds roughly to the strength of "diabetes" or "hypertension" individually — a confounder of that strength acting on top of all the existing adjustments would shift the CI to include the null.

Limitations

1. **The E-value framework assumes a single unmeasured confounder** with a specified joint association strength. Multiple weaker unmeasured confounders could in principle combine to explain the association even if no single confounder reaches the E-value threshold.

2. **The E-value addresses only confounding**, not other biases (selection bias, measurement error, time-varying effects). The complementary Bayesian probabilistic bias analysis (Appendix S3) addresses some of these directly.

3. **The Chinn (2000) approximation** for converting standardized continuous effects to risk-ratio scale is approximate; more refined methods exist (Mathur et al. 2018) but yield qualitatively similar results.

Implications

The E-values for the all-cause mortality (2.92) and PhenoAge Acceleration (3.04) findings indicate substantial robustness to unmeasured confounding. Combined with the Bayesian probabilistic bias analysis (Appendix S3) and the negative-control analysis (Appendix S10), this triangulation of methods provides convergent evidence that the OD–biological-aging association is not easily explained by residual confounding in plausible scenarios.

External replication in non-NHANES cohorts (pre-registered in Appendix S7, OSF DOI 10.17605/OSF.IO/JK2G7) remains the strongest test of confounding robustness, particularly in cohorts with measured neurodegeneration biomarkers, frailty assessments, and ApoE-ε4 genotype.

References

- VanderWeele TJ, Ding P. (2017). Sensitivity analysis in observational research: introducing the E-value. Annals of Internal Medicine 167(4): 268–274.
- Chinn S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. Statistics in Medicine 19(22): 3127–3131.
- Mathur MB, et al. (2018). Web site and R package for computing E-values. Epidemiology 29(5): e45–e47.

Companion deposits

- Code (Zenodo): https://doi.org/10.5281/zenodo.20059714
- Manuscript archive (OSF): https://doi.org/10.17605/OSF.IO/7Y5FU

- Pre-registered external replication (OSF): https://doi.org/10.17605/OSF.IO/JK2G7

Supplementary Appendix S13: Clinical Translation Analyses

Purpose

This appendix provides clinical-relevance analyses to inform potential future clinical deployment of the Pocket Smell Test, including: (i) absolute risk gradients across clinical subgroups; (ii) number-needed-to-screen by population characteristics; (iii) diagnostic test characteristics at multiple OD cutoffs; and (iv) head-to-head comparison with established clinical risk markers. These analyses extend the manuscript's primary §3.8 translational analyses with subgroup-targeted clinical metrics.

Important framing: These analyses identify the population subgroups where olfactory testing might have the most plausible clinical value — they do *not* establish that screening should be recommended. Decision curve analysis in §3.8 of the manuscript shows that OD ≥4 carries prognostic information correlated with biological aging but does not improve clinical risk stratification beyond standard ASCVD variables at conventional decision thresholds. The analyses below identify populations where the prognostic gradient is largest, but external replication in non-NHANES cohorts is required before any clinical deployment can be considered.

Methods

The analytic sample includes 3,512 NHANES 2013–2014 adults aged ≥40 years with mortality follow-up through December 31, 2019 (median follow-up 5.92 years; 373 deaths). Five-year cumulative mortality risk was estimated by simple Kaplan-Meier estimation within each subgroup. Number-needed-to-screen (NNS) is defined as 1/absolute-risk-difference between OD ≥4 and no-OD strata at 5 years. Diagnostic test characteristics use a 5-year follow-up window for the binary outcome.

Results

1. Absolute risk gradient by clinical subgroup

Subgroup	n	OD ≥4 prevalence	5-yr mortality, OD ≥4	5-yr mortality, no OD	Absolute risk increase	NNS
Overall (≥40 yr)	3,512	8.9%	31.8%	10.2%	21.6 pp	5
No chronic disease	938	5.2%	13.1%	2.9%	10.2 pp	10
Any chronic disease	2,574	10.2%	35.3%	13.0%	22.3 pp	4

Subgroup	n	OD ≥4 prevalence	5-yr mortality, OD ≥4	5-yr mortality, no OD	Absolute risk increase	NNS
Diabetes	761	12.1%	38.8%	15.6%	23.2 pp	4
Hypertension	2,335	10.1%	36.0%	13.4%	22.5 pp	4
Cancer history	473	12.7%	53.8%	23.1%	30.7 pp	3
Age 40-59	1,838	4.1%	2.7%	4.5%	-1.9 pp	–
Age 60-69	877	8.4%	8.1%	8.8%	-0.7 pp	–
Age 70+	797	20.3%	55.5%	28.0%	27.5 pp	4
Diabetes + age 60+	473	15.9%	47.3%	17.9%	29.4 pp	3
No comorbidities + age 60+	212	11.3%	27.4%	9.9%	17.5 pp	6

Interpretation: The absolute risk gradient associated with OD ≥4 varies substantially by clinical subgroup. In adults under 60 years of age, OD ≥4 does not identify excess mortality risk (point estimates negative or near zero). In adults aged ≥70 or with cardiometabolic disease, OD ≥4 identifies a high-mortality population (NNS = 3-4). The cancer-history subgroup shows the steepest gradient (NNS = 3, ARR = 30.7 pp), although this estimate is based on only 473 individuals and requires external replication.

These findings suggest that, if olfactory testing were ever to be considered clinically, it would be most informative in older adults with established cardiometabolic disease — the population where the absolute prognostic gradient is largest and the NNS is smallest. In younger adults without comorbidities, the prognostic gradient is absent or unclear, and screening would not be expected to improve clinical risk stratification.

2. Four-stratum mortality table (the central clinical observation)

Stratum	n	Events	5-yr mortality	Risk diff vs reference	RR vs reference
Reference: No OD + No chronic disease	889	24	2.9%	—	1.0
OD ≥4 + No chronic disease	49	6	13.1%	+10.2 pp	4.5×
No OD + Chronic disease	2,312	257	13.0%	+10.2 pp	4.5×

Stratum	n	Events	5-yr mortality	Risk diff vs reference	RR vs reference
OD ≥4 + Chronic disease	262	86	35.3%	+32.4 pp	12.2×

Interpretation: The combination of OD ≥4 and chronic cardiometabolic disease identifies a population with 12-fold higher 5-year mortality than the reference healthy population. The two factors appear approximately additive in absolute risk terms (each adds ~10 pp), but combined they identify a population with substantial 5-year mortality (>35%) that may warrant attention even without further intervention research. The OD ≥4 + No chronic disease stratum (n = 49) is the smallest stratum and produces the most uncertain estimate, but its 13.1% 5-year risk is consistent with the chronic-disease-only stratum, suggesting that OD ≥4 carries similar prognostic information as having any of {diabetes, hypertension, cancer history}.

3. Diagnostic test characteristics at clinical thresholds

3a. Whole sample (n = 3,512; 5-year all-cause mortality, prevalence 10.6%)

Test	Sensitivity	Specificity	PPV	NPV	PLR	NLR
OD ≥3 (NHANES standard)	39.4%	84.6%	23.4%	92.2%	2.57	0.72
OD ≥4 (proposed cutoff)	24.7%	93.0%	29.6%	91.2%	3.54	0.81
OD ≥5 (severe OD)	12.9%	97.4%	36.6%	90.4%	4.87	0.89

3b. Chronic-disease subgroup (n = 2,574; 5-year mortality, higher prevalence)

Test	Sensitivity	Specificity	PPV	NPV	PLR
OD ≥3	39.4%	83.0%	26.3%	89.9%	2.32
OD ≥4	25.1%	92.1%	32.8%	88.9%	3.18
OD ≥5	12.8%	97.1%	40.4%	87.9%	4.40

Interpretation: At the proposed OD ≥4 cutoff in the whole sample, a positive test increases the probability of 5-year mortality from 10.6% (baseline) to 29.6% (PPV) — a 2.8-fold increase in post-test probability. The positive likelihood ratio (PLR = 3.54) places OD ≥4 in the range that biostatisticians commonly classify as “moderate” diagnostic utility (PLR 2-5). In the chronic-disease subgroup, the PLR is similar (3.18) but the higher pre-test probability of mortality means the absolute post-test probability of mortality (PPV = 32.8%) is itself clinically substantial.

The trade-off across cutoffs is clear: OD ≥3 has higher sensitivity but lower PPV; OD ≥5 has higher PPV but only catches 12.9% of eventual deaths. The OD ≥4 cutoff approximates an inflection point in this trade-off.

4. Comparison with established clinical risk markers (Cox HRs, age + sex adjusted only)

Marker	n	% positive	Hazard ratio	95% CI	p-value
OD ≥4	3,512	8.9%	**1.83**	1.43-2.34	<0.001
HbA1c ≥6.5% (diabetes range)	3,512	15.3%	1.31	1.03-1.68	0.029
Creatinine ≥1.3 mg/dL (mildly elevated)	3,512	8.1%	1.94	1.51-2.50	<0.001
SBP ≥140 mmHg (stage 1 HTN)	3,512	21.6%	1.05	0.84-1.31	0.661
Albumin <3.5 g/dL (hypoalbuminemia)	3,512	1.6%	3.73	2.44-5.70	<0.001
NLR ≥3.0 (elevated inflammation)	3,512	17.6%	1.62	1.30-2.02	<0.001

Interpretation: With minimal adjustment (age and sex only), OD ≥4 carries comparable mortality prognostic information to mildly elevated creatinine — a marker that is universally measured in routine clinical care. OD ≥4 has a stronger HR than HbA1c ≥6.5% (the diabetes diagnostic threshold) and stronger than stage-1 hypertension. It has a weaker HR than hypoalbuminemia (a marker of severe illness with low population prevalence) but a stronger HR than elevated NLR. Importantly, OD ≥4 uses no laboratory testing — only an 8-item olfactory examination requiring approximately 5 minutes.

This comparison is informative for understanding the *type* of prognostic information OD ≥4 carries: it is in the range of routinely-measured laboratory markers of organ-system stress (creatinine, NLR, albumin), not in the range of "tells you almost nothing" (SBP ≥140 with adjustment alone).

5. Clinical-deployment scenario quantification

Among older adults (≥60 years) with chronic cardiometabolic disease (n = 1,462 in this sample, representing approximately 50 million U.S. adults nationally):

- **OD ≥4 prevalence in this population**: 14.5%
- **5-year mortality risk if OD ≥4**: 42.4%

- **5-year mortality risk if no OD**: 18.4%
- **Absolute risk increase**: 24.0 percentage points
- **Number-needed-to-screen to identify 1 high-risk individual**: 4

Per 100,000 chronic-disease 60+ adults screened, OD ≥4 would identify approximately 14,500 high-risk individuals; the population-attributable difference in 5-year deaths between OD ≥4 and no-OD strata is approximately 3,500. This is a research scenario quantification, not a clinical recommendation.

Caveats: 1. These quantifications assume the NHANES association generalizes; external replication in ARIC/Health ABC/BOSS is pre-registered (Appendix S7). 2. NNS is a screening-yield metric; it does not establish that screening *prevents* deaths — this would require an intervention study with confirmed outcome modifiability. 3. The decision curve analysis in §3.8 of the manuscript shows negative net benefit at conventional ASCVD-style decision thresholds, indicating that OD ≥4 augmentation of standard risk calculators is not clinically justified at conventional cutoffs. 4. The clinical deployment scenarios above are framed as "research signal" demonstrations, not screening recommendations.

Synthesis: clinical implications

These analyses position the OD ≥4 cutoff as carrying **clinically meaningful prognostic information** in a specific population — older adults with established cardiometabolic disease — comparable in HR magnitude to mildly elevated creatinine and stronger than several routinely-used cardiovascular risk markers. The PPV at OD ≥4 in the whole sample is approximately 30%; in the chronic-disease subgroup it is approximately 33%.

What this *does* suggest: - OD ≥4 is biologically meaningful and prognostically informative in plausibly-targeted populations - The chronic-cardiometabolic-disease subgroup is the clinically relevant population for any future deployment research - The OD ≥4 cutoff (rather than NHANES standard ≥3) is the more clinically informative threshold

What this *does not* establish: - That screening should be recommended (decision curve analysis at conventional thresholds: negative net benefit) - That OD ≥4 identifies modifiable risk (no intervention study has tested this) - That the population-level deployment quantifications generalize beyond U.S. NHANES (external replication required)

External replication in non-NHANES cohorts (ARIC, Health ABC, BOSS) and prospective intervention studies are the necessary next steps before any clinical-deployment recommendations can be considered.

Companion deposits

- Code (Zenodo): https://doi.org/10.5281/zenodo.20059714
- Manuscript archive (OSF): https://doi.org/10.17605/OSF.IO/7Y5FU
- Pre-registered external replication (OSF): https://doi.org/10.17605/OSF.IO/JK2G7

Supplementary Table S1

Table S1. Characteristics of analytic sample by measured olfactory dysfunction status using the **conventional NHANES OD cutoff (≥3 wrong)**, NHANES 2013–2014 adults aged ≥40 years (n = 3,519).

Characteristic	No OD (n = 2,889)	OD ≥3 wrong (n = 630)	p
DEMOGRAPHICS			
Age, years	56.7 (11.1)	65.0 (12.4)	<0.001
Female, %	53.6	43.8	<0.001
Non-Hispanic White, %	72.3	64.3	0.029
<HS education, %	13.5	27.0	<0.001
Poverty-income ratio	3.20 (1.64)	2.66 (1.53)	<0.001
LIFESTYLE			
Current smoker, %	18.2	14.3	0.039
Heavy alcohol, %	7.0	3.0	<0.001
Meets PA guidelines, %	50.9	38.5	<0.001
CLINICAL			
BMI, kg/m^2	29.5 (6.8)	29.6 (7.1)	0.887
Waist circumference, cm	101.5 (15.6)	104.0 (16.5)	<0.001
Systolic BP, mmHg	125.1 (17.5)	130.5 (19.9)	<0.001
Diabetes, %	16.7	22.3	0.001
Hypertension, %	62.2	73.1	<0.001
Cancer history, %	14.8	23.9	0.001
Sinonasal disease, %	89.8	90.4	0.632
BIOMARKERS			
NLR (ratio)	2.32 (1.21)	2.52 (1.62)	0.004
Platelets, ×10^9/L	232 (59)	218 (60)	<0.001
BIOLOGICAL AGING			
KDM-BA, years	66.9 (24.1)	80.3 (29.6)	<0.001

Characteristic	No OD (n = 2,889)	OD ≥3 wrong (n = 630)	p
KDM-BA Acceleration	−3.27 (20.95)	+0.96 (26.99)	<0.001
Homeostatic Dysregulation	2.42 (1.16)	2.89 (1.20)	<0.001
Modified PhenoAge, years	58.1 (14.2)	68.5 (16.6)	<0.001
PhenoAge Acceleration	−0.74 (8.46)	+0.80 (9.88)	<0.001

Note: Survey-weighted means (SD) for continuous variables; n (weighted %) for categorical variables. P-values from design-adjusted t-test or Rao-Scott χ^2. This table preserves the conventional NHANES OD definition (≥3 of 8 odors incorrectly identified) for descriptive comparability with prior NHANES literature. The headline biological-aging and mortality analyses in the main manuscript use the moderate-to-severe cutoff (≥4 wrong) — see Table 1 for the corresponding characteristics.

Nature Portfolio Reporting Summary

Manuscript: Olfactory testing identifies accelerated biological aging and mortality risk: NHANES 2013–2014 **Author:** Ceyhun Olcan **Corresponding author:** ceyhun.olcan.27@dartmouth.edu

This Reporting Summary follows the Nature Portfolio reporting requirements for life sciences research. For the official PDF form, see https://www.nature.com/documents/nr-reporting-summary.pdf — this document provides the answers in markdown format for transcription into the official form at submission.

Statistics

For all statistical analyses, confirm that the following items are present in the relevant location.

n/a / Confirmed

Sample size: ✓ Confirmed - The exact sample size (N = 3,519 cross-sectional; N = 2,983 Cox-eligible; N = 373 deaths) is reported in Methods §2.1, Results §3.1, Results §3.7, and Figure 1 (STROBE flow diagram).

Data exclusions: ✓ Confirmed - Exclusions are reported in Figure 1 (STROBE flow): n = 6,360 aged <40 excluded; n = 107 not in chemosensory exam frame; n = 189 incomplete PST. All exclusions were pre-specified before data analysis. No data were excluded post hoc based on outcome values.

Replication: ✓ Confirmed - Internal split-sample replication across 20 random PSU partitions is reported in Results §3.4 and Methods §2.5; mean HR 1.73 (range 1.11–2.43); 100% of splits showed HR > 1.0. - Methodological replication of biological aging clock construction in NHANES 2011–2012 is reported in Supplementary Appendix S1.B. - Empirical NLR-for-CRP substitution validation in NHANES 2017–March 2020 (n = 5,474; Pearson r = 0.997) is reported in Supplementary Appendix S1.A. - External replication in independent cohorts (ARIC, Health ABC, Beaver Dam Offspring Study) is identified as future work in Discussion §4.8.

Randomization: n/a (observational study; no random assignment of exposure) - The 1:1 propensity score matching procedure is described in Methods §2.5, with covariate balance assessment shown in Figure 4.

Blinding: n/a (observational study using public-use de-identified data)

Reporting for specific materials, systems, and methods

We require information from authors about some types of materials, experimental systems, and methods used in many studies.

Materials & experimental systems

Item	Involved?
Antibodies	n/a
Eukaryotic cell lines	n/a
Palaeontology and archaeology	n/a
Animals and other organisms	n/a
Human research participants	**YES**
Clinical data	**YES**
Dual use research of concern	n/a

Methods

Method	Used?
ChIP-seq	n/a
Flow cytometry	n/a
MRI-based neuroimaging	n/a

Human research participants

Policy information about studies involving human research participants

Population characteristics: - Source: U.S. NHANES 2013–2014, the only NHANES cycle with a fully implemented measured olfactory examination. - Inclusion: Adults aged ≥40 years with a valid 8-item Pocket Smell Test result (N = 3,519). - Demographics: 53.1% female; mean age 58.0 years (SD 12.5); 71.2% non-Hispanic White, 10.6% non-Hispanic Black, 4.8% non-Hispanic

Asian, 6.9% Mexican American, 6.5% other Hispanic. - Setting: U.S. non-institutionalized civilian population, with NHANES sample weights for nationally-representative inference.

Recruitment: - NHANES uses a complex multistage probability sample of the non-institutionalized U.S. population. Detailed recruitment procedures are documented in NHANES protocols available at https://wwwn.cdc.gov/nchs/nhanes/. The 2013–2014 cycle achieved an 8-item Pocket Smell Test completion rate of 94.9% among eligible aged ≥40 year participants.

Ethics oversight: - NHANES is conducted under NCHS Research Ethics Review Board approval (Protocol #2011-17, with continuing review through 2024). All participants provided written informed consent. - The present secondary analysis used fully de-identified public-use data and was exempt from additional institutional review board oversight.

Clinical data

Policy information about clinical studies

This is **not a clinical trial** and trial registration is not applicable. The work is a secondary analysis of an existing public surveillance dataset.

Clinical trial registration: Not applicable - The analysis is observational and based on existing public-use NHANES data.

Study protocol: - The retrospective analysis plan is publicly archived on the Open Science Framework: [OSF DOI to be assigned] - The complete analytic pipeline is publicly archived on Zenodo: [Zenodo DOI: 10.5281/zenodo.20059714]

Data collection: - All data were collected by NCHS as part of the standard NHANES program. The author had no role in data collection. The author had access to the de-identified public-use data files only.

Outcomes: - Primary biological aging outcome: Modified PhenoAge Acceleration (residual after regressing modified PhenoAge on chronological age). - Primary mortality outcome: All-cause mortality through December 31, 2019, ascertained from the NCHS Public-Use Linked Mortality File. - Cause-specific outcomes: Cardiovascular mortality (UCOD_LEADING==1, ICD-10 codes I00–I09, I11, I13, I20–I51) and cancer mortality (UCOD_LEADING==2, ICD-10 codes C00–C97). - All outcomes were specified before primary model interpretation.

Software and code

Policy information about software and code:

Data collection: - Software: NHANES data are collected by NCHS using their standard protocols and software. The analyst did not participate in data collection.

Data analysis: - Python 3.12 with NumPy 1.26, SciPy 1.11, pandas 2.1, scikit-learn 1.3, matplotlib 3.7. - Analytic code is publicly available at https://github.com/ceyhunolcan/od-aging-nhanes

(Zenodo DOI: 10.5281/zenodo.20059714). - All custom implementations (Pocket Smell Test scoring, KDM-BA, Homeostatic Dysregulation Mahalanobis distance, modified PhenoAge with NLR-for-CRP substitution, Geskus Fine-Gray IPCW reformulation, multi-mediator natural direct/indirect effect decomposition with PSU-clustered bootstrap, Bayesian probabilistic bias analysis) are documented in the source code with line-level comments.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data availability

Policy information about availability of data:

Confirmation that data are available: ✓ All data underlying the analyses are publicly available without restriction.

- **NHANES 2013–2014** source data: https://wwwn.cdc.gov/nchs/nhanes/ (cycle G files: DEMO_H, BIOPRO_H, BMX_H, BPQ_H, BPX_H, CBC_H, CSQ_H, DIQ_H, DPQ_H, GHB_H, HDL_H, MCQ_H, PAQ_H, SMQ_H, TCHOL_H, CSX_H)
- **Public-Use Linked Mortality File**: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm (file: NHANES_2013_2014_MORT_2019_PUBLIC.dat)
- **NHANES 2017–March 2020** (used for NLR-for-CRP validation): https://wwwn.cdc.gov/nchs/nhanes/
- **Derived analytic datasets**: `analytic_with_clocks.csv` (n = 3,519 with computed biological aging clocks) and `analytic_with_mort.csv` (n = 3,519 with mortality follow-up) are provided as supplementary materials and archived on Zenodo with the analytic code.
- **Source Data file**: `source_data.xlsx` containing the underlying data behind every figure is provided as a supplementary file.

No restricted-use NHANES data were used. No personally-identifying information is contained in the analytic dataset.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Field	Selected?
Life sciences	✓ **Selected**
Behavioural & social sciences	(not selected — closer fit but observational

Field	Selected?
	epidemiology with biomarkers fits better under life sciences)
Ecological, evolutionary & environmental sciences	n/a

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size: - N = 3,519 was determined by the available NHANES 2013–2014 sample with a complete 8-item Pocket Smell Test. No formal a priori power calculation was performed because the sample size was fixed by the underlying public dataset. Post hoc power assessment: with 302 events in the Cox-eligible cohort and an OD ≥4 prevalence of 6.3% weighted, the analysis had >85% power to detect HR ≥ 1.5 at $\alpha = 0.05$.

Data exclusions: - All exclusions were pre-specified and are documented in Figure 1 (STROBE flow diagram). Aged <40 (n = 6,360 excluded) per protocol; not in chemosensory exam frame (n = 107); incomplete Pocket Smell Test (n = 189). No exclusions were made post hoc based on outcome values.

Replication: - Internal split-sample replication across 20 random PSU partitions: 100% of splits yielded HR > 1.0; mean HR = 1.73 (range 1.11–2.43). - Methodological replication of biological aging clock construction in NHANES 2011–2012 (Supplementary Appendix S1.B): cycle-to-cycle stability of clock distributions confirmed. - Empirical NLR-for-CRP substitution validation in NHANES 2017–March 2020 (Supplementary Appendix S1.A): Pearson r = 0.997, Bland-Altman bias +0.34 years.

Randomization: - Not applicable (observational study; no randomized assignment of exposure). Confounder control via covariate adjustment, propensity score matching, and inverse probability weighting.

Blinding: - Not applicable (observational study using publicly-available de-identified data).

End of Nature Portfolio Reporting Summary

This document summarizes the responses required by the Nature Portfolio Reporting Summary form.

For submission, transcribe these responses into the official Nature Reporting Summary PDF form available at https://www.nature.com/documents/nr-reporting-summary.pdf and submit alongside the manuscript.
