Brain, Behavior, and Immunity 137 (2026) 106806

Contents lists available at ScienceDirect

Brain Behavior and Immunity

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

Strategies for analyzing multiple inflammatory biomarkers: Impacts on the strength and replicability of findings in the MIDUS study

Phoebe H. Lam [a,*], Gregory E. Miller [b,c]

[a] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
[b] Department of Psychology, Northwestern University, Evanston, IL, USA
[c] Institute for Policy Research, Northwestern University, Evanston, IL, USA

A B S T R A C T

Circulating biomarkers of low-grade inflammation are increasingly used in behavioral medicine and related disciplines as subclinical indicators of health problems. However, strategies for analyzing these biomarkers vary widely, and it is unclear whether such analytical variability affects the assumed inflammation-health association. Using the Midlife in The United States (MIDUS) cohorts as large case examples, we examined whether three analytical strategies – (1) using raw vs. log-transformed data; (2) analyzing biomarkers individually vs. forming composites; and (3) if the latter, how they were constructed– affected the magnitude and replicability of inflammation-health associations. The Core ($N = 1,201$) and Refresher ($N = 709$) cohorts included data on 8 inflammatory biomarkers and 8 health problems. A subset of Core participants ($N = 643$) had longitudinal data about 9 years later. Three composites were constructed: reflective (via factor analysis), formative (via PLS-SEM), and agnostic (standardized mean scores) scales using both raw and log-transformed data. These variations yielded 24 operationalizations of inflammation (2 transformation methods × 12 scoring approaches [8 individual biomarkers + 4 composites]). We then evaluated each operationalization based on (a) its cross-sectional and longitudinal associations with health outcomes, in terms of statistical significance and effect size; and (b) the replicability of the cross-sectional associations across the two samples, in terms of significance agreement and effect size consistency. In MIDUS, operationalizations based on log-transformed (vs. raw) biomarker data yielded stronger and more replicable associations with health outcomes, and so did composite-based operationalizations (vs. individual biomarkers). Patterns were evident in cross-sectional models, and to a lesser extent, in longitudinal models. These findings are not intended to prescribe a single best-practice for analyzing inflammatory biomarkers, but rather to illustrate how analytical decisions can shape estimated inflammation-health links, and thus, how well these biomarkers function as subclinical disease indicators within a given sample.

1. Introduction

Interest in biomarkers of low-grade inflammation has increased in behavioral medicine and related disciplines. Because their elevated levels predict morbidity of many health conditions (Furman et al., 2019; Guo et al., 2013; Li et al., 2017), these biomarkers are often used in observational studies as indicators of subclinical disease or potential mediators linking psychosocial factors to health. However, analytical practices vary: some analyze biomarkers individually, whereas others create composites. Recent work has used psychometric methods to evaluate composites (Moriarity et al., 2021; Egnot et al., 2018) or examined how such choices influenced links with psychological factors (Rengasamy et al., 2023). Yet, few have evaluated analytical practices against the rationale underlying the use of inflammatory biomarkers in much of behavioral medicine: as replicable indicators of health problems. Using two Midlife in the United States (MIDUS) cohorts as large case examples, this study examined how (a) using raw vs. log-transformed data, (b) analyzing biomarkers individually vs. forming composites; and (c) the composites construction method affected the magnitude and replicability of inflammation-health associations.

1.1. Inflammatory biomarkers as preclinical indicators of disease

Inflammation is a defense mechanism against pathogens and tissue injury. Detection of these triggers immune cells to release cytokines, such as interleukin (IL)-1b, tumor necrosis factor-alpha (TNF-α), and IL-6, which coordinate immune-cell recruitment, activation, proliferation, and differentiation (Abbas et al., 2014). This response also involves up-regulation of endothelial adhesion molecules, such as E-selectin and Intercellular Adhesion Molecule-1 (ICAM), facilitating immune cell extravasation into target tissues (Haraldsen et al., 1996). IL-6 also stimulates production of acute-phase proteins, such as C-Reactive Protein (CRP) and fibrinogen, which mark pathogens for destruction (Pepys and Hirschfield, 2003) and support coagulation to limit spreading (De Vries et al., 2020), respectively. Together, cytokines, adhesion molecules, and acute-phase proteins coordinate the acute phase of

* Corresponding author at: 4825 Frew St, Suite 354E, Pittsburgh, PA 15213, USA.
E-mail address: phoebelam@cmu.edu (P.H. Lam).

https://doi.org/10.1016/j.bbi.2026.106806
Received 17 October 2025; Received in revised form 24 April 2026; Accepted 10 May 2026
Available online 12 May 2026

inflammation.

Persistent stimuli or failure in regulating inflammatory activities, however, can lead to chronic low grade, "non-resolving" inflammation, which may contribute to various health problems (Nathan and Ding, 2010). Animal studies provide causal evidence for inflammation's role in disease-relevant processes (Chae et al., 2010; Hoogland et al., 2015; McNeill et al., 2010; Nakano-Narusawa, et al., 2910). In humans, over 100 prospective studies found that circulating levels of proteins described above forecasted disease morbidity, functional declines, and premature death (Furman et al., 2019; Guo et al., 2013; Danesh et al., 2000; Fragão-Marques et al., 2025; Kaptoge et al., 2010; Lai et al., 2017; Marcos-Pérez et al., 2020; Michels et al., 2021). Mendelian Randomization studies further suggest that at least some inflammatory biomarkers may be causally related to diseases (Bottigliengo et al., 2022; Vandebergh et al., 2022; Wei et al., 2023). Together, this work provides an empirical basis for using inflammatory biomarkers as proxies for early pathological processes.

1.2. Immunoassay advances and analytical variability

With the advent of multiplex immunoassays, studies increasingly assess multiple inflammatory biomarkers: in the literature on childhood stress and inflammation (Chiang et al., 2022), the number of markers assessed per study increased from one before 2000, up to 4 in 2000–2009, and up to 10 in 2010–2021. Yet, researchers differed in how they analyzed them, including decisions about transformation and aggregation.

Raw* vs. *Transformed Data. One source of variability is whether to analyze biomarkers in their native units to facilitate biological and clinical interpretation or normalize their distributions (e.g., log transformation) to reduce complications in statistical analyses.[1] Our first aim, thus, was to evaluate the impacts of using native units vs. log-transformation.

Individual biomarker approach. Another source of variability concerns whether to analyze individual biomarkers or as composites. Analyzing biomarkers separately is biologically sensible as proteins can originate from different tissues and serve distinct functions (Medzhitov, 2021; Pradeu et al., 2024). It is also conceptually appealing for examining specificity links between specific biomarkers and particular outcomes (Dinarello, 2011; Picard et al., 2003; Schett et al., 2013). In humans, however, such hypotheses may be difficult to evaluate (Miller, 2009), particularly with circulating biomarkers as they are released across time, tissues, and stimuli, obscuring the granularity likely needed for inferences about specificity. Testing specificity hypotheses also requires isolating unique associations often by statistically adjusting for other biomarkers; yet, many biomarkers are meaningfully correlated because they are co-induced and signal through overlapping or convergent pathways (Haddad, 2002). Residualization may thus remove variance intrinsic to the inflammatory process under study and yield net values with unclear biological interpretation. Finally, analyzing biomarkers individually increases false discoveries and reduces replicability, especially when multiple-test corrections are inconsistently applied (García-Pérez, 2023; Ioannidis, 2014; Rubin, 2024; Unger et al., 2025).

Composite approach. Biomarker composites lower the risks of false discovery and are used in 'omics fields via dimension-reduction methods suited for high-dimensional data (Misra et al., 2019). However, their suitability when only a small number of inflammatory biomarkers are available is unclear. Considering the conceptual assumptions underlying different composite approaches may help. In *reflective* models, a latent construct (e.g., low-grade inflammation) is assumed to cause variation in its observed indicators (e.g., IL-6), which are, in theory, correlated manifestations of the same underlying process. By contrast, *formative* models assume that indicators define the construct, with each indicator contributing potentially uniquely, so covariation among indicators is not assumed nor required (Edwards and Bagozzi, 2000).

How do measurement models align with the biology of inflammation? In early inflammatory responses, infected or damaged tissues release cytokines and chemokines that recruit and activate leukocytes (Vestweber, 2012). At this stage, commonly studied biomarkers like IL-1ß and TNF-α are initiating components of the inflammation response. At this stage of the response the direction of influence runs from indicators to construct, consistent with a *formative* model. However, as the inflammatory response progresses, particularly in non-resolving lesions, leukocyte activity drives coordinated changes in many of the biomarkers typically studied. In other words, at this later stage, an underlying inflammatory state likely drives biomarker release, so the direction of influence runs from the construct to indicators, consistent with a *reflective* model. Both models, therefore, may be applicable but reflect different temporal stages. Studies have examined reflective models via factor analyses (Moriarity et al., 2021; Egnot et al., 2018; Koukkunen et al., 2001; Kristono et al., 2020; Pripp and Stanišić, 2014; Tziakas et al., 2007), but none have applied formative models (PLS-SEM) to inflammatory biomarkers or compared the two.

Finally, a third approach to composite formation is to remain "agnostic" about underlying structure by standardizing biomarkers into common units and averaging the resulting values. Unlike the formative and reflective approaches, this approach does not weight individual markers and may minimize the risk of sample- or outcome-specific weights that can limit replicability. Yet, by weighting all biomarkers equally, the agnostic approach could obscure differences in biomarker functions or the strength of their links with health outcomes.

1.3. Evaluation criteria

Recent work examined psychometric features (e.g., factor structure, internal reliability) (Moriarity et al., 2021; Egnot et al., 2018) and how analytic decisions impact links with psychological factors (Rengasamy et al., 2023; Moriarity et al., 2021). But when analytical evaluation is embedded within hypothesis testing, it is unclear whether an approach that supports a hypothesis is better or simply aligns with expectations. We propose that an approach's utility should also depend on how well it serves the conceptual basis underlying its use. In observational studies of behavioral medicine, inflammatory biomarkers are often measured as subclinical indicators of health problems. Accordingly, at least one criterion should be whether it yields *strong* and *replicable* links with those health outcomes.

Two studies have examined how analytical decisions affect inflammation-health links (Moriarity et al., 2021; Kristono et al., 2020). One found that an IL-6-IL-8 composite predicted cardiovascular events more strongly than either biomarker alone, or a larger composite of six biomarkers (Kristono et al., 2020). The other examined the reverse direction, entering seven health problems simultaneously to predict inflammatory biomarkers; only diabetes showed consistent links with inflammation, and the model including the CRP-fibrinogen and IL-8-IL-10 composites vs. the model with only individual biomarkers yielded similar conclusions (Moriarity et al., 2021). No studies evaluated the second criterion proposed here: whether analytical decisions impact the replicability of inflammation-health links.

[1] When fit as outcomes in linear models, right-skewness can lead to non-normal residuals (see Supplementary Fig. S1). When used as predictors, isolated values can exert excessive leverage, even in models that do not assume normal residuals (e.g., logistic regression; see Supplementary Figs. S1B). Robust methods can preserve native units (Moriarity, 2022), but simulation studies suggest they may underperform for some parameters under high skewness, especially in small samples (Lai, 2018). This simulation defined skewness of 2 as high, whereas biomarkers like IL-6 can exhibit skewness exceeding 20 (e.g., in MIDUS).

Table 1
Descriptive statistics and bivariate spearman correlations among inflammatory biomarkers.

MIDUS 2 & 3	Mean (SD) in SI units	Mean (SD) in molar units	1	2	3	4	5	6	7	8
1. CRP	3.03 (4.78); 4.09 (5.76)	131.4 (207.6); 175.9 (252)	**0.59***	0.55*	0.53*	0.26*	0.09*	0.03	0.29*	0.37*
2. Fibrinogen	348.6 (86.7); 379.18 (87.3)	9.14 (2.28); 9.91 (2.25)	0.53*	**0.50***	0.45*	0.20*	0.09*	0.07	0.25*	0.24*
3. IL-6	1.09 (1.00); 1.44 (1.40)	0.05 (0.05); 0.07 (0.07)	0.52*	0.41*	**0.58***	0.41*	0.22*	0.13*	0.24*	0.27*
4. TNF-α	2.17 (0.73); 2.30 (0.80)	0.13 (0.04); 0.13 (0.05)	0.23*	0.11*	0.35*	**0.45***	0.34*	0.27*	0.21*	0.19*
5. IL-10	0.28 (0.21); 0.30 (0.24)	0.01 (0.01); 0.02 (0.01)	0.08*	0.05	0.15*	0.34*	**0.49***	0.20*	0.12*	0.12*
6. IL-8	13.4 (6.35); 14.6 (8.32)	1.6 (0.76); 1.71 (0.98)	0.02	0.05	0.11*	0.20*	0.20*	**0.55***	0.08*	0.09*
7. ICAM	288.0 (115); 231.0 (84.2)	5.22 (2.08); 4.20 (1.55)	0.19*	0.13*	0.18*	0.28*	0.13*	0.07*	**0.47***	0.43*
8. E-Selectin	43.30 (22.67); 37.10 (19.6)	0.67 (0.35); 0.58 (0.32)	0.15*	0.11*	0.15*	0.09*	0.10*	0.08*	0.05	**0.69***
Refresher	**Mean (SD) in SI units**	**Mean (SD) in molar units**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. CRP	2.96 (5.18)	128.6 (224.8)	–							
2. Fibrinogen	343.4 (73.75)	9.01 (1.93)	0.57*	–						
3. IL-6	1.03 (0.98)	0.05 (0.05)	0.56*	0.46*	–					
4. TNF-α	2.14 (0.82)	0.12 (0.05)	0.23*	0.18*	0.37*	–				
5. IL-10	0.29 (0.23)	0.02 (0.01)	0.04	0.01	0.14*	0.39*	–			
6. IL-8	12.03 (5.78)	1.43 (0.69)	0.05	0.13*	0.15*	0.23*	0.11*	–		
7. ICAM	257.3 (96.71)	4.66 (1.75)	0.16*	0.14*	0.14*	0.23*	0.18*	0.02	–	
8. E-Selectin	40.61 (18.95)	0.63 (0.29)	0.24*	0.18*	0.28*	0.29*	0.18*	0.05	0.37*	–

Note. *$p < 0.05$. For each inflammatory biomarker, means and standard deviations are reported for MIDUS 2 (first value), MIDUS 3 (second value), and Refresher cohort. Biomarkers are reported in both mass-based SI units and molar units. CRP is reported in μg/mL; fibrinogen is reported in mg/dL; IL-6, TNF-α, IL-10, and IL-8 are reported in pg/mL; E-selectin and ICAM are reported in ng/mL. For molar reporting, CRP, ICAM, and E-selectin are expressed in nanomolar (nM); fibrinogen is expressed in micromolar (μM); and IL-6, TNF-α, IL-10, and IL-8 are expressed in picomolar (pM). Spearman correlations for MIDUS 2 are presented below the diagonal, whereas those from MIDUS 3 are presented above the diagonal. Diagonal values (bolded) reflect longitudinal associations from MIDUS 2 to MIDUS 3 for each marker. CRP = C-reactive protein; IL = Interleukin; TNF-α = Tumor necrosis factor- alpha; ICAM = Intercellular Adhesion Molecule-1.

1.4. The current investigation

This study examined how transformation and aggregation decisions affected the magnitude and replicability of the inflammation-health associations using MIDUS Core and Refresher samples. These cohorts provide large, probability-based samples of midlife to older adults, when health problems are more prevalent, but findings may not generalize across populations, biomarker panels, or study designs. Thus, rather than prescribing a single best practice, we use these cohorts as case examples to show that analytical decisions can impact how well inflammatory biomarkers proxy as preclinical indicators of health problems.

2. Method

2.1. Participants and procedure

We conducted analyses on the public datasets MIDUS 2, MIDUS 3, and Refresher cohorts. The core MIDUS cohort (N = 7,108) was recruited in 1995–1996 via random-digit dialing. Inclusion criteria include non-institutionalized, English-speaking adults living in the U.S., and aged 25 to 74 (Brim et al., 2020). MIDUS 2 and MIDUS 3 were follow-up assessments conducted about 9 and 18 years later, with 4,963 and 3,294 participants, respectively. About 1,255 of MIDUS 2 participants completed the Biomarker Project, with 747 returning at MIDUS 3. The Refresher sample (N = 3,577) was recruited in 2011–2014 (Ryff et al., 2017) using the same eligibility criteria, and 863 Refresher participants completed the Biomarker Project (Weinstein et al., 2019). The

Table 2
Descriptive statistics and bivariate odds ratios among health conditions.

MIDUS 2 & 3	MIDUS 2 N (%)	MIDUS 3 N (%)	New cases N (%)	1	2	3	4	5	6	7	8
1. Cancer	169 (14%)	171 (25%)	107 (18%)	**60.29***	1.25	1.10	1.12	1.22	2.39*	2.03*	1.17
2. Arthritis	513 (43%)	361 (53%)	161 (37%)	2.46*	**7.01***	1.44	1.81*	1.76*	1.61*	2.00*	1.22
3. Respiratory	174 (14%)	139 (21%)	57 (10%)	1.61*	1.88*	**149.89***	1.66*	1.40	1.63*	1.47	0.77
4. Blood	231 (19%)	187 (28%)	103 (18%)	1.84*	1.82*	1.65*	**12.38***	1.23	1.88*	1.45*	0.82
5. Endocrine	284 (23%)	243 (36%)	119 (22%)	1.71*	1.59*	1.20	1.30	**33.2***	1.53*	1.30	1.58*
6. Cardiovascular	303 (25%)	226 (33%)	116 (22%)	1.36	1.84*	1.54*	1.60*	1.51*	**19.6***	1.49*	1.03
7. Gastrointestinal	270 (22%)	289 (43%)	179 (33%)	1.65*	2.31*	1.52*	1.71*	1.78*	1.73*	**13.71***	1.40*
8. Metabolic	417 (34%)	232 (34%)	96 (21%)	0.99	1.19	1.03	0.60*	1.60*	1.07	1.16	**6.23***
Refresher	**RefresherN (%)**			**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. Cancer	123 (14%)			–							
2. Arthritis	319 (37%)			1.61*	–						
3. Respiratory	181 (21%)			2.71*	1.37	–					
4. Blood	220 (26%)			3.11*	1.49*	1.73*	–				
5. Endocrine	210 (24%)			2.30*	1.88*	1.8*	2.22*	–			
6. Cardiovascular	189 (22%)			2.76*	0.95	2.09*	1.75*	2.01*	–		
7. Gastrointestinal	245 (28%)			1.45*	0.91	1.12	1.40	2.24*	2.21*	–	
8. Metabolic	295 (34%)			1.68*	1.33	0.70	2.53*	1.87*	1.17	1.04	–

Note. *$p < 0.05$. For each health condition, frequency (N) and prevalence (%) are reported. Crude odds ratios for MIDUS 2 are presented below the diagonal, whereas those from MIDUS 3 are presented above the diagonal. Diagonal values (bolded) reflect longitudinal associations from MIDUS 2 to MIDUS 3 for each health condition. The proportion of new cases was computed as the prevalence of a given health condition among participants without that condition at MIDUS 2.

current analyses were restricted to participants with inflammation data, yielding N's = 1,201 (MIDUS 2), 709 (Refresher), and 643 (MIDUS 3). The MIDUS 2 sample had a mean age of 54.6 years (SD = 11.7; range = 34–84), was 57% female, and 77% White; the Refresher sample had a mean age of 50.8 years (SD = 13.4; range = 25–76), was 50% female, and 81% White; and the MIDUS 3 sample had a mean age of 61.0 years (SD = 9.78; range = 43–90), was 55% female, and 90% White.

2.2. Measures

Inflammatory biomarkers. Fasting blood samples were collected in the early morning (between 6 and 8:30am; minimizing diurnal variation) into sodium citrate and serum separator tubes, centrifuged at 4°C for 15 min at 2000–3000 rpm and for 20 min at 2000–3000 rpm, respectively. Citrated plasma and sera were stored at −60°C to −80°C. Eight inflammatory biomarkers were assayed: CRP and fibrinogen from citrated plasma, IL-6, IL-8, IL-10, TNF-α, soluble E-Selectin, and soluble ICAM from sera. Supplementary Materials provide biological and methodological details. Table 1 provides descriptive statistics of inflammatory biomarkers, and Supplementary Table S1 provides these statistics stratified by health problem endorsement.

Health problems. Participants self-reported whether they had 14 health problems, grouped into 7 categories: cardiovascular (heart disease, heart murmur, transient ischemic attack/stroke), respiratory (asthma, emphysema/chronic obstructive pulmonary disease, or tuberculosis), endocrine (diabetes or thyroid disease), gastrointestinal (peptic ulcer disease or colon polyp), blood-related (blood clots, anemia, or other blood disorders), cancer, and arthritis. In MIDUS 2 and Refresher, each category was dummy coded such that "no" was coded as 0, and "yes" or "borderline" were coded as 1. In MIDUS 3, the response scale changed to "yes, diagnosed", "yes, undiagnosed", "borderline, diagnosed", and "borderline, undiagnosed", in addition to "no". Thus, responses were similarly dummy coded such that "no" was coded as 0, whereas "yes" and "borderline" coded as 1, regardless of diagnosis status. In addition, a metabolic category was created based on metabolic syndrome status, defined using the International Diabetes Federation (IDF) adult criteria: central obesity (waist circumference ≥ 94 cm for men, ≥80 cm for women) plus at least two additional risk factors (triglycerides ≥ 150 mg/dL, low HDL cholesterol, elevated blood pressure, or fasting glucose ≥ 100 mg/dL). Table 2 presents prevalence, new cases at follow-up, and links among health problems.

2.3. Analytical approach

Supplemental Materials provide detailed descriptions of the analytical approach, which we briefly summarize here. Prior to data analyses, distributions were inspected and outliers were identified (±3.5 *SD* from mean) and removed (Fig. S2; MIDUS 2: removed 30–57 [2–5%]; Refresher: removed 22–28 [3%]).

Composite formation. In Phase 1, we constructed inflammation composites using three approaches: reflective, formative, and agnostic. For the reflective approach,[2] we conducted parallel analyses to determine the number of factors, followed by an exploratory factor analysis (EFA) in MIDUS 2 and then a confirmatory factor analysis (CFA) in Refresher. Factor analyses were estimated using maximum likelihood estimation with robust standard errors (MLR) and Geomin oblique rotation. Model fit was evaluated based on conventional standards (Hu and Bentler, 1999). Regression factor scores were extracted from the CFA models and used in Phase 2 analyses.

[2] We did not adopt the previously published factor structure derived from MIDUS datasets (Moriarity et al., 2021) because of potential outlier influence (as summarized in Supplementary Table S2 and Figs. S2-S3) and because factors with only two indicators may be unstable across replications (p.23; Brown, 2015).

In the formative approach, we constructed composites using PLS-SEM (Sarstedt et al., 2022). Latent inflammation composites were formed by estimating outer weights (i.e., the contribution of each biomarker to the composite), optimized to maximize variance explained in the eight health outcomes. Statistical inference was based on bootstrapping. We formed composites based on cross-sectional (MIDUS 2 and Refresher) and longitudinal (MIDUS 2 inflammation predicting MIDUS 3 health, controlling for the corresponding MIDUS 2 health outcome) inner models. Cross-sectional composite scores were extracted for Phase 2 analyses (Refresher did not have follow-up data). Composite-level outliers were excluded (0.2 to 1.3%); retaining these values did not alter the aggregated pattern of results or overall conclusions. Finally, in the agnostic approach, we averaged standardized (Z-scored) of the eight inflammatory biomarkers.

Evaluation. In Phase 2, we evaluated how analytical approach impacted the magnitude and replicability of inflammation-health associations using logistic regressions. Predictors were standardized, and links were expressed in odds ratios (ORs; exponentiated coefficients). Magnitude was summarized by (a) statistical significance: the proportion of significant associations using conventional ($p < 0.05$) and Bonferroni-corrected (0.05/8 health outcomes = $p < 0.00625$) thresholds; and (b) effect size: the geometric mean OR across the 8 health problems.

Replicability was assessed by comparing results from MIDUS 2 vs. Refresher, summarized by (a) significance agreement: whether analyses yielded the same inference (significant vs. not) across the two samples; and (b) effect size consistency: Pearson's *r* and intraclass correlation coefficients (ICCs; two-way random-effects model assessing consistency), characterizing correspondence in log ORs across the two samples.

3. Results

Table 1 presents bivariate correlations among inflammatory biomarkers. Correlation magnitudes varied. Across cohorts, CRP, fibrinogen, and IL-6 consistently showed the strongest correlations (ρ's = 0.41–0.57), whereas correlations involving CRP with IL-8 or IL-10 (0.02–0.09), fibrinogen with IL-10 (ρ's = 0.01–0.09), as well as ICAM with IL-8 (0.02–0.08) were weakest. All biomarkers showed moderate to strong temporal stability over the 9-year period (ρ's = 0.45–0.69). Supplementary Materials provide full details on model specifications and more granular operationalization-specific findings. Readers can also select biomarkers and health outcomes to visualize how inclusion or exclusion of specific biomarkers or health outcomes changes certain estimates using an interactive Shiny App (https://phoebelam.shinyapps.io/shinymidus/).

3.1. Construction of reflective composites

As summarized in Table S3, EFA in MIDUS 2 suggested consistent structures across raw and log-transformed data: CRP, fibrinogen, and IL-6 loaded onto one; TNF-α, IL-10, IL-8, and ICAM loaded onto another. E-selectin did not load adequately onto either. CFA confirmed this structure for both raw and transformed data. Model fit was acceptable to good (CFI = 0.93, RMSEA = 0.057–0.079, SRMR = 0.041–0.047), except for TLI (0.88–0.89). Fit improved after adding the residual covariance between IL-6 and TNF-α (CFI = 0.95–0.96; TLI = 0.91–0.92; RMSEA = 0.049–0.063; SRMR = 0.037–0.040). Reliability was higher for log-transformed data ($\omega_{factor1} = 0.76$, $\omega_{factor2} = 0.47$) than raw data ($\omega_{factor1} = 0.47$, $\omega_{factor2} = 0.13$). Factor scores in each model were extracted.

Supplementary analyses. We further examined the role of sample size in replicability of the factor loadings by drawing 1,500 random subsamples per sample size (100–500) from MIDUS 2 with replacement, re-estimated CFAs, and computed the proportion of loadings that fell within the 95% CI from the full *Refresher* sample. The proportion of

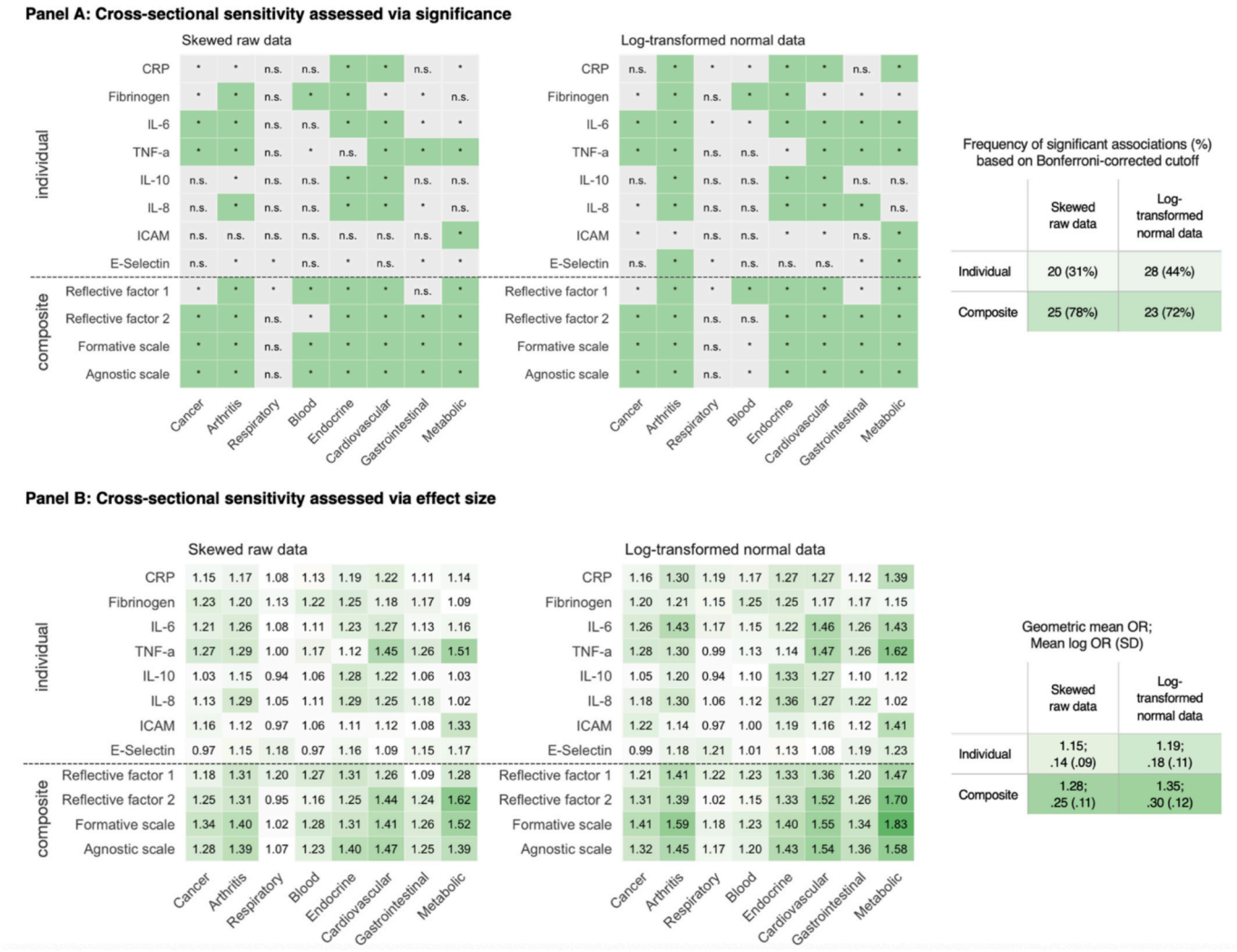

Fig. 1. Cross-sectional inflammation-health associations stratified by analytical approach. Panel A shows the sensitivity of individual biomarker and composite scores to health problems based on statistical significance. Color indicates Bonferroni-significance (green = significant; grey = not significant), and text denotes results at conventional threshold (* = significant, and "n.s." = not significant). "Conv sig" = conventional threshold (0.05) and "Bonf sig" = Bonferroni threshold (0.00625). Panel B shows sensitivity based on effect size, with color indicating the magnitude of odds ratio (OR; grey = smallest; green = largest) and text showing the OR value. Analytical N varies across associations due to pairwise data availability; thus, ORs of the same magnitude may differ in statistical significance.

loadings falling within the Refresher CIs approached acceptable levels (at least 75%) only when N ≥ 400 (Fig. S6).

3.2. Construction of formative composites

Formative composites were constructed using PL-SEM. As summarized in Table S4, in MIDUS 2, skewness handling mattered. Using raw skewed data, IL-6, TNF-α, and IL-8 contributed significantly to the composite; using log-transformed data, these same biomarkers still contributed, but so did CRP and ICAM. In the Refresher sample, IL-6, TNF-α, and IL-8, CRP, fibrinogen, and E-selectin all contributed to the composite for both raw and log-transformed data in Refresher.

We then repeated analyses based on its associations with health problems at MIDUS 3, controlling for the corresponding baseline health status at MIDUS 2. Again, no stable formative structure emerged: Only IL-6, TNF-α, and ICAM contributed significantly to the composite when using log-transformed data, whereas only fibrinogen, TNF-α, and E-Selectin contributed significantly when using raw skewed data. Formative composite scores were extracted per sample.

3.3. Comparisons of sensitivity: cross-sectional associations

As shown in Fig. 1 (operationalization × health outcome) and Fig. S4A (operationalization averaged across health outcomes), we then conducted head-to-head comparisons of approaches. All health problems, except respiratory problems, were significantly associated with at least one operationalization of inflammation, after Bonferroni-corrections.

Raw skewed data* vs. log-*transformation. As summarized in Table 3, log transformation yielded a higher proportion of significant associations than raw skewed data (75% vs. 67% at conventional; 53% vs. 47% with Bonferroni correction). Moreover, log transformation also

Table 3
Summary of sensitivity of inflammation–health associations by analytic approaches.

Cross-sectional		Effect size				Significance			
Marker	Transformation	OR	Log OR	CIL	CIU	N / total	%	CIL	CIU
Individual	Skewed raw	1.15	0.14	0.12	0.16	20/64	31%	20%	43%
Individual	Log-normal	1.19	0.18	0.15	0.20	28/64	44%	32%	56%
Composite	Skewed raw	1.28	0.25	0.21	0.28	25/32	78%	64%	92%
Composite	Log-normal	1.35	0.30	0.26	0.35	23/32	72%	56%	87%
Individual	–	1.17	0.16	0.14	0.18	48/128	38%	29%	46%
Composite	–	1.32	0.27	0.24	0.30	48/64	75%	64%	86%
–	Skewed raw	1.19	0.18	0.15	0.20	45/96	47%	37%	57%
–	Log-normal	1.25	0.22	0.19	0.24	51/96	53%	43%	63%
Longitudinal		**Effect size**				**Significance**			
Marker	**Transformation**	**OR**	**Log OR**	**CIL**	**CIU**	**N / total**	**%**	**CIL**	**CIU**
Individual	Skewed raw	1.06	0.06	0.03	0.09	2/64	3%	0%	7%
Individual	Log-normal	1.08	0.08	0.04	0.11	5/64	8%	1%	14%
Composite	Skewed raw	1.12	0.11	0.07	0.15	5/32	16%	3%	28%
Composite	Log-normal	1.14	0.13	0.08	0.17	5/32	16%	3%	28%
Individual	–	1.07	0.07	0.05	0.09	7/128	5%	2%	9%
Composite	–	1.13	0.12	0.09	0.15	10/64	16%	7%	25%
–	Skewed raw	1.08	0.08	0.05	0.10	7/96	7%	2%	12%
–	Log-normal	1.10	0.09	0.07	0.12	10/96	10%	4%	17%

Note. Sensitivity was based on both average effect size (geometric mean of odds ratios [OR] and average log OR) and the proportion of Bonferroni-corrected significant associations out of total tested within each category. CIL and CIU indicate confidence interval lower and upper bounds. A dash (–) indicates that results are collapsed across the omitted analytic decision (i.e., averaged across levels of that factor). Specifically, a dash in the transformation column indicates results collapsed across skewed raw and log-normal specifications, whereas a dash in the marker column indicates results collapsed across individual and composite biomarkers.

showed larger average effect size (OR = 1.25, log OR = 0.22 [0.19, 0.24]) compared to raw data (OR = 1.19, log OR = 0.18 [0.15, 0.20]).

Individual biomarkers* vs. *composites. As summarized in Table 3, compared to individual markers, composites yielded a greater proportion of significant associations based on conventional (88% vs. 63%) and Bonferroni-corrected (75% vs. 38%) thresholds. Composites also yielded larger average effect sizes: 1 SD increase in composite score was associated with an average 32% increase in the odds of health problems (OR = 1.32, Log OR = 0.27 [0.24, 0.30]), compared to 17% for individual markers (OR = 1.17, Log OR = 0.16 [0.14, 0.18]).[3]

Because many individual markers were non-significant correlates of health, we recomputed mean ORs among only significant associations (i.e., $p < 0.05$). Composite scores still showed larger average effect size (OR = 1.35 vs. 1.24). Moreover, because formative scales are designed to maximize prediction of outcomes, we further excluded formative scales in these re-computations. Composites still showed larger average effect size (OR = 1.33 vs. 1.24).

3.4. Comparisons of sensitivity: longitudinal associations

Fig. 2 shows comparisons of significance and effect sizes using longitudinal models. Inflammation predicted increases in risk for endocrine and cardiovascular problems nine years later, controlling for the corresponding baseline status.

Both individual markers and composites yielded few Bonferroni-adjusted significant associations (0–25%) and small average effect size (OR's = 1.02–1.18). Nonetheless, on average, composites showed more significant associations (16% vs. 5%) and larger average effect size (1.13 vs. 1.07). The effect size difference remained when recomputed among only conventionally significant associations (composite OR = 1.34, individual OR = 1.26). Note that this relative difference was small when formative scales were excluded. Furthermore, the differences by skewness handling method were small both in terms of proportion of Bonferroni-significant associations (raw = 7% vs. logged = 10%) and average effect size (raw = 1.08 vs. logged = 1.10).

[3] Composite-level outliers were excluded (0.2% to 1.3%); retaining these values did not alter the aggregated pattern of results or overall conclusions.

3.5. Comparisons of replicability

Fig. 3 (operationalization × health outcome) and Fig. S4B (operationalization averaged across health outcomes) show significance agreement and effect size consistency.

Raw skewed data* vs. log-*transformation. As summarized in Table 4 and Fig. 4, using Bonferroni-corrected thresholds, log transformation yielded a higher proportion of replicated *significant* associations (n = 35, 36%) compared to raw data (n = 23, 24%), and more consistent estimates of effect sizes (*ICCs* = 0.47 vs. 0.27; r = 0.58 vs. 0.33). When considering replicability of *non-significant* associations, the approach did not matter (logged: n = 37, 39%; raw: n = 38, 40%).

Individual biomarkers* vs. *composites. As summarized in Table 4 and Fig. 4, using Bonferroni-corrected thresholds, composites yielded more replicable *significant* associations (n = 31, 48%) compared to individual markers (n = 27, 21%). By contrast, individual markers yielded more replicable *non-significant* associations (n = 64, 50%) than composites (n = 11, 17%).

Because more total tests were done for individual markers than composites (128 vs. 64), we recomputed proportions out of only initially significant associations. Patterns remained: Composites showed more replicated *significant* links (74% vs. 30%), whereas individual markers showed more replicated *non-significant* links (70% vs. 26%). The same pattern remained after further removing formative scales: composites still showed more replicated *significant* links (68% vs. 30%), individual markers showed more replicated *non-significant* links (70% vs. 32%).

In terms of effect size consistency across sample, there was substantial variability in ICCs and *r's* across approaches, ranging from poor to acceptable (*ICC* = −0.08–0.68; r = −0.13–0.82). Composites yielded effect sizes that were relatively more consistent, falling in the fair range (*ICC* = 0.51, r = 0.62), than individual markers which fell in the poor range (*ICC* = 0.30, r = 0.38).

3.6. Exploratory analyses

We repeated analyses controlling for age, sex, and race. Several patterns emerged, depicted in Fig. S5. First, across all operationalizations, fewer associations were Bonferroni-significant after covariate adjustment (50% vs. 28%). Second, differences between composites and

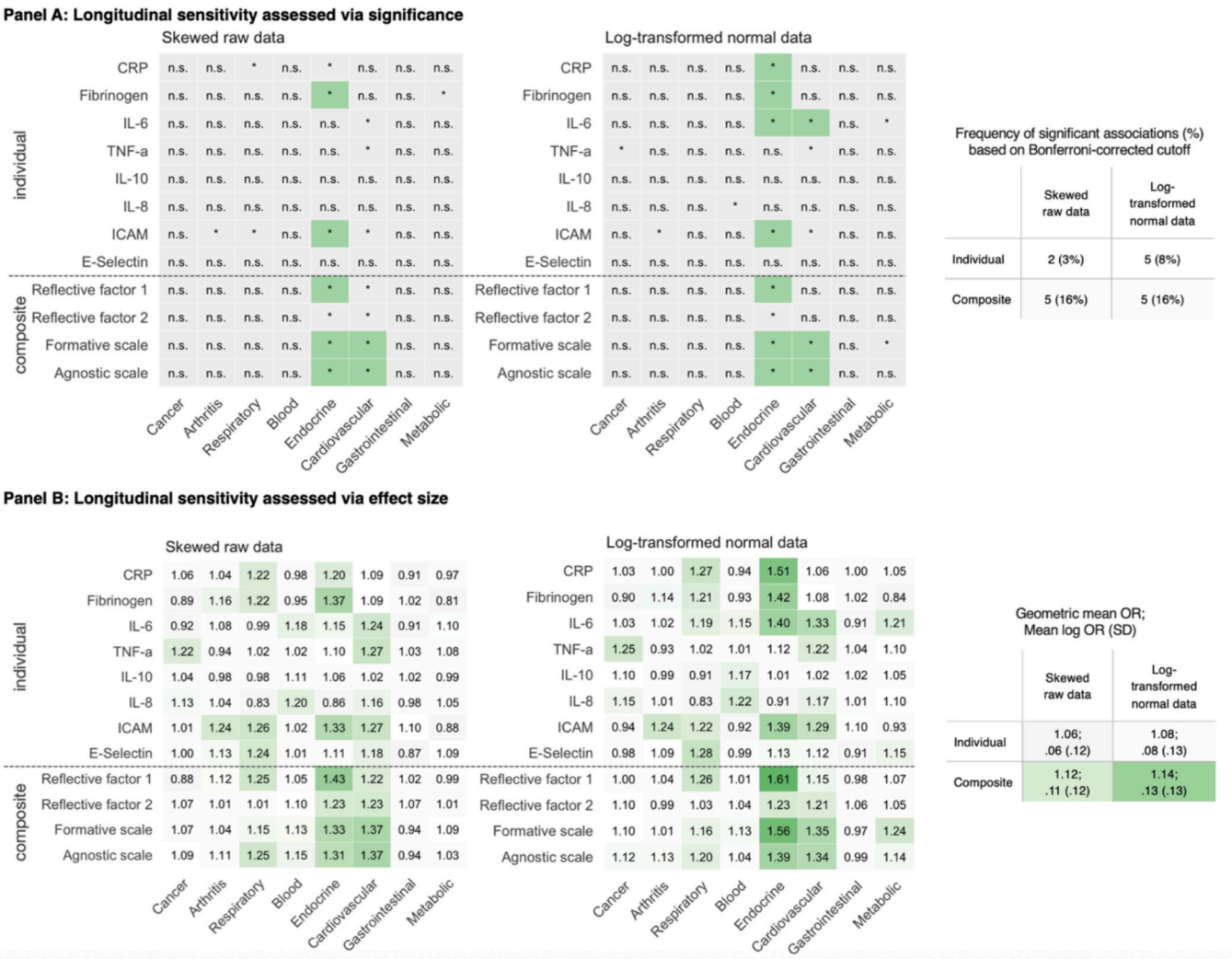

Fig. 2. Longitudinal inflammation-health associations stratified by analytical approach. Panel A shows the sensitivity of individual biomarker and composite scores to health problems based on statistical significance. Color indicates Bonferroni-significance (green = significant; grey = not significant), and text denotes results at conventional threshold (* = significant, and "n.s." = not significant). "Conv sig" = conventional threshold (0.05) and "Bonf sig" = Bonferroni threshold (0.00625). Panel B shows sensitivity based on effect size, with color indicating odds ratio magnitude (OR; grey = smallest; green = largest) and text showing the OR value. Predictors were standardized such that OR reflect change in odds of health problems per 1-SD increase in inflammation. Analytical N varied across associations due to pairwise data availability; thus, ORs of the same magnitude may differ in statistical significance.

individual markers remained but were smaller: composites showed a higher proportion of Bonferroni-corrected significant associations (39% vs. 22%), larger average effect sizes (1.23 vs. 1.13), more replicated significant associations (22% vs. 11%), and greater consistency in effect sizes ($r = 0.86$ vs. 0.63, $ICC = 0.78$ vs. 0.49). Third, the difference between using transformed and raw data also remained but were smaller: transformation showed a higher proportion of Bonferroni-corrected significant associations (32% vs. 23%), larger average effect sizes (1.18 vs. 1.14), more replicated significant associations (16% vs. 14%), and greater consistency in effect sizes ($r = 0.79$ vs. 0.66; $ICC = 0.69$ vs. 51). Finally, as above, differences by analytical approach were limited for longitudinal models.

Because dichotomizing biomarkers is a common and clinically informative practice that also allows retaining extreme values, we explored the impact of analyzing individual biomarkers continuously vs. categorically. These analyses were limited to CRP (>3 mg/L; Pearson et al., 2003), fibrinogen (>400 mg/dL; Gasparoto et al., 2021), IL-6 (≥3.19 pg/mL; Harris et al., 1999), and TNF-α (>2.5 pg/mL; Khudiakova et al., 2023), for which there was at least one published threshold derived from generally healthy adults. As depicted in Fig. S7, continuous indicators yielded more Bonferroni-significant associations with higher replication rates across cohorts, whereas categorical indicators produced larger but poorly replicating effect sizes.

4. Discussion

Across two MIDUS cohorts, log-transformed (vs. raw) data and composites (vs. individual biomarkers) yielded stronger and more replicable links with health outcomes. These patterns were observed in cross-sectional and, to much lesser extent, adjusted and longitudinal models. As illustrated in the MIDUS study, how inflammation is operationalized may affect both the strength and replicability of its links to health outcomes and thus impact how well these biomarkers fulfill their intended roles as preclinical indicators of health problems.

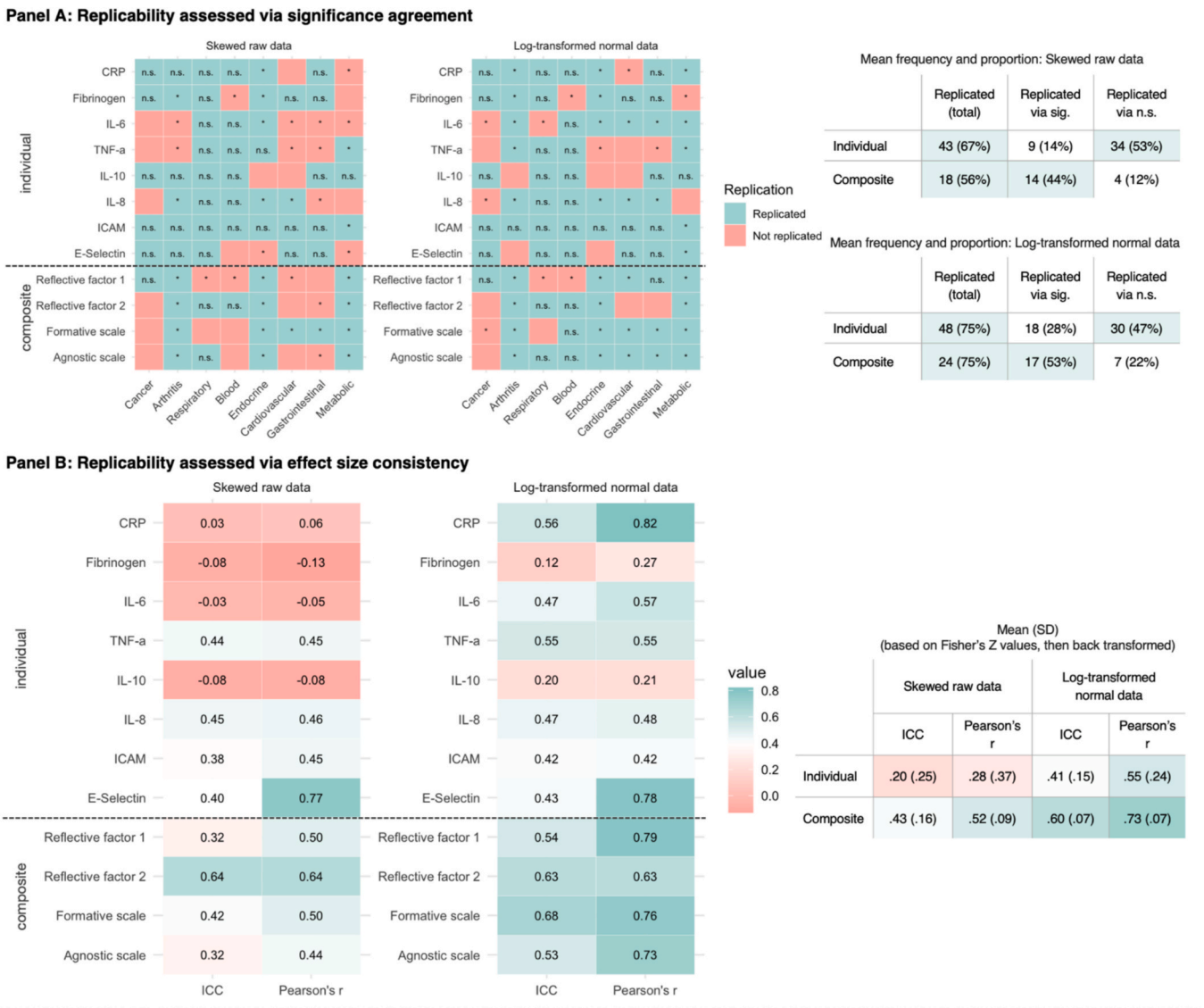

Fig. 3. Replicability of the inflammation-health link stratified by analytical approach. Panel A shows replication based on significance agreement. Associations were considered replicated if both samples yielded the same conclusion (both significant or both not) using Bonferroni-corrected p-values. Green cells indicate replication; red cells indicate non-replication. Within green cells (replicated), * indicates replicated *significant* associations, while "n.s." indicates replicated *non-significant* associations. Red cells with * reflect replication if using $p < 0.05$ threshold. Tables summarize the frequency and proportion of replicated links by scoring approach and skewness handling (i.e., % out of row totals: 64 for individual and 32 for composite). Panel B shows effect size consistency using intraclass correlation coefficients (ICCs) and Pearson's *r* values. The table summarizes the mean (SD) ICC and *r* values (based on Fisher's Z values, then back transformed) by scoring approach and skewness handling.

4.1. Can reflective and formative composites be reliably formed?

A consistent two-factor structure emerged with CRP, IL-6, and fibrinogen as a factor, and IL-10, TNF-α, IL-8, and ICAM as another. This structure replicated across MIDUS cohorts for both raw and log-transformed data.[4] Yet, only the log-transformed CRP-IL-6-fibrinogen factor, which was also observed in two other prior studies (Egnot et al., 2018; Koukkunen et al., 2001), showed acceptable internal reliability. These proteins shared variance attributable to a common latent factor, potentially because they are involved in the hepatic acute-phase signaling pathway mediated by JAK/STAT3 (Ngwa et al., 2022; Albrecht et al., 2007). By contrast, although TNF-α, IL-10, IL-8, I-CAM formed a stable factor across cohorts, this factor exhibited poor internal reliability. Finally, factor loadings were within acceptable replicability ranges only when N ≥ 400, highlighting practical concerns for forming reflective constructs.

There was no replicable evidence supporting a formative conceptualization across MIDUS cohorts. Formative models are most applicable if biomarkers represent distinct molecular activities that jointly give rise to an inflammatory state, as may occur during earlier phases of the inflammatory response. Circulating biomarkers, or the ones measured here, may not adequately assess these earlier inflammatory processes that are relatively localized, trigger-specific, and cell-dependent (Medzhitov, 2021; Pradeu et al., 2024). Formative model instability may also be due to methodological issues, including mixed evidence on PL-SEM's adequacy for constructing stable and unbiased formative composites (Aguirre-Urreta and Marakas, 2008) and the lower sensitivity of distal, self-reported binary outcomes, compared to continuous,

[4] The observed factor structure differed from that reported in prior work using the same datasets, which identified CRP and fibrinogen on one factor and IL-8 and IL-10 on the other (Moriarity et al., 2021), potentially due to differences in how extreme values were handled (see Supplementary Materials).

Table 4
Summary of replicability of inflammation–health associations by analytic approaches.

Consistency of effect size		Average intra-class correlation			Average Pearson's r		
Marker	Transformation	ICC	CIL	CIU	*r*	CIL	CIU
Individual	Skewed raw	0.20	0.02	0.37	0.28	0.01	0.51
Individual	Log-normal	0.41	0.30	0.51	0.55	0.36	0.69
Composite	Skewed raw	0.43	0.26	0.58	0.52	0.43	0.60
Composite	Log-normal	0.60	0.52	0.67	0.73	0.66	0.79
Individual	–	0.31	0.20	0.42	0.42	0.25	0.57
Composite	–	0.52	0.42	0.61	0.64	0.54	0.72
–	Skewed raw	0.28	0.14	0.42	0.37	0.18	0.53
–	Log-normal	0.48	0.39	0.56	0.62	0.49	0.72

Agreement of significance		**Replication via sig.**				**Replication via. n.s.**			
Marker	**Transformation**	**N / total**	**%**	**CIL**	**CIU**	**N / total**	**%**	**CIL**	**CIU**
Individual	Skewed raw	9 / 64	14%	6%	23%	34 / 64	53%	41%	65%
Individual	Log-normal	18 / 64	28%	17%	39%	30 / 64	47%	35%	59%
Composite	Skewed raw	14 / 32	44%	27%	61%	4 / 32	12%	1%	24%
Composite	Log-normal	17 / 32	53%	36%	70%	7 / 32	22%	8%	36%
Individual	–	27 / 128	21%	14%	28%	64 / 128	50%	41%	59%
Composite	–	31 / 64	48%	36%	61%	11 / 64	17%	8%	26%
–	Skewed raw	23 / 96	24%	15%	32%	38 / 96	40%	30%	49%
–	Log-normal	35 / 96	36%	27%	46%	37 / 96	39%	29%	48%

Note. Replicability was evaluated using effect size consistency (mean intraclass correlation coefficients [ICC] and Pearson's r for Log ORs between MIDUS 2 and Refresher) and significance agreement (concordant significant [sig] or non-significant [n.s.] results in both samples). Significance was determined based on Bonferroni corrected threshold ($p < 0.00625$). CIL and CIU indicate confidence interval lower and upper bounds. A dash (–) indicates that results were collapsed across the omitted analytic decision (i.e., averaged across levels of that factor). Specifically, a dash in the transformation column indicates results collapsed across skewed raw and log-normal specifications, whereas a dash in the marker column indicates results collapsed across individual and composite biomarkers.

more proximal, or objective outcomes.

4.2. Can analytic choices shape the strength and replicability of inflammation–health associations?

In the MIDUS study, on average, log-transforming (vs. raw) data and composites (vs. individual biomarkers) yielded stronger and more reliable associations with health problems. Notably, composites showed greater replication of *significant* associations, whereas individual biomarkers showed greater replication of *null* associations, suggesting that the larger composite effect sizes were unlikely to be driven by random noise. Except for log-transformed IL-6, individual biomarkers either yielded smaller effect sizes or observed significant associations that largely did not replicate. Moreover, analyzing individual biomarkers categorically based on published thresholds did not improve performance, but rather yielded fewer significant associations and larger, yet poorly replicating, effect sizes. This pattern may be because dichotomizing continuous variables inflates effect sizes and false positive rates (MacCallum et al., 2002).

In the MIDUS study, among composite types, the formative scales were designed to, and indeed, produced the strongest links with health outcomes, but they should not be considered unless the instability issues mentioned above are resolved. (The general pattern of results described above remained when formative scales were removed). The log-transformed agnostic composite, constructed without weighing indicators, yielded the next strongest associations with health outcomes and showed good replicability for significant links. This may be because agnostic composites avoid sample-specific weights linked with any one biomarker or that individual contribution of each marker is less important relative to their joint elevation. However, there are likely limits to the usefulness of aggregating agnostically. The inflammatory biomarkers assessed in MIDUS is a relatively small panel; future research will be necessary to identify conditions when agnostic vs. weighted composites are preferable.

4.3. Additional patterns

Two additional patterns emerged. The first pattern relates to whether the inflammation-health link in MIDUS is characterized by *specificity* (different biomarkers show associations with different conditions) or by *generality* (biomarkers tend to show associations across multiple outcomes). If specificity were present, one or a few biomarker(s) would reliably predict (or show stronger effect sizes with) only certain conditions, whereas other biomarker(s) would predict different conditions. We did not observe such patterns here. Cross-sectionally, arthritis, cardiovascular, metabolic, and endocrine conditions were each predicted by at least five log-transformed biomarkers, and respiratory conditions showed uniformly null associations. Seeming cases of specificity (e.g., fibrinogen with blood-related conditions; IL-6 and TNF-α with cancer) did not replicate across MIDUS cohorts, and occurred only for the outcomes with the lowest prevalence (cancer = 14%, blood = 19% vs. 23% to 43% for all others), so may be better explained by limited prevalence than biological specificity. These findings do not rule out specificity but suggest that it was not detectable with *the current design using MIDUS data*. In particular, basal, circulating biomarkers are themselves low in specificity and sensitivity to discrete disease processes, as they are pooled signals released across time, tissues, and stimuli, obscuring the granularity likely necessary for inferences about specificity. Future research can use in-vitro measurements of inflammatory processes, quasi-experiments that compare clinical vs. well-matched groups, and multivariable approaches or classification-based methods to test specificity.

Second, few longitudinal links between baseline inflammation and health outcomes nine years later emerged after adjusting for baseline status; significant links were limited to endocrine and cardiovascular conditions. This differs from prior larger prospective studies with greater incidents at follow-up (thousands to tens of thousands vs. average ~120 here; e.g., Danesh et al., 2008; Qin et al., 2025; Zhu et al., 2022). Unlike our approach of adjusting for baseline status, these studies also excluded those with baseline disease and modeled longitudinal links among initially healthy participants. Still, the possibility that circulatory inflammatory markers do not predict changes in health problems cannot be ruled out, so future work should evaluate this in larger prospective samples.

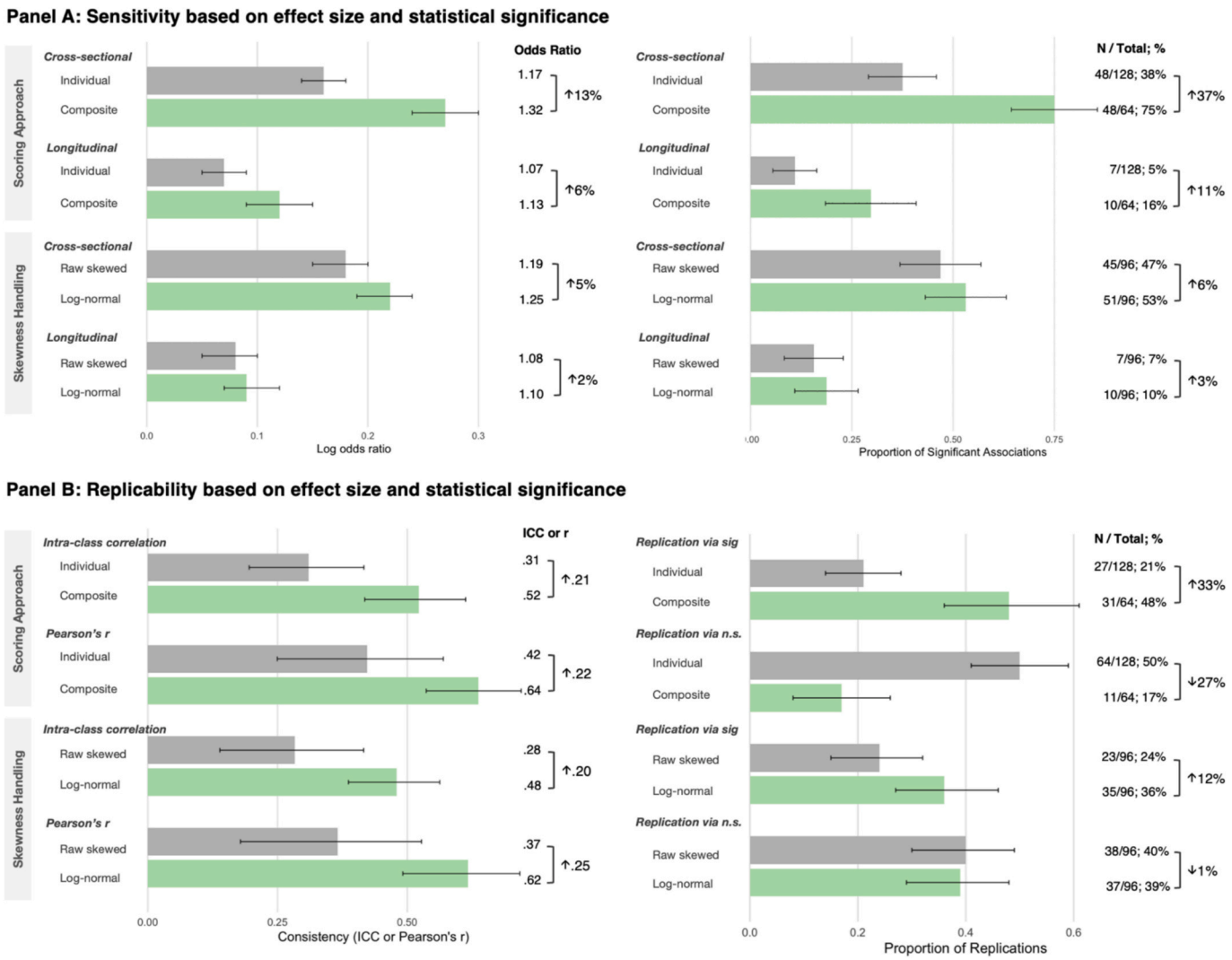

Fig. 4. Summary of sensitivity and replicability of inflammation–health associations by analytic approaches. Panel A summarizes average effect size (geometric mean odds ratios [OR]) and the proportion of significant associations. Relative percent differences in OR and absolute differences in proportions were computed by analytical approach. Panel B summarizes replicability based on effect size consistency (mean intraclass correlation coefficients [ICC] and mean Pearson's r) and the proportion of replicated associations (the number of associations that replicated by being significant [sig] or non-significant [n.s.] in both samples). Absolute differences in ICC/r and replication proportions were computed by analytical approach. In all cases, significance was defined with Bonferroni correction applied.

4.4. Implications and future directions

Taken together, this study found that log-transformed composites yielded stronger and more replicable associations in the MIDUS study. Along with psychometric evaluations, if indicator weighing is desired and if analytical sample size is at least 400, the log-transformed CRP-fibrinogen-IL-6 factor is a reasonable operationalization, but if weighing is not preferred or analytical sample size is limited, a log-transformed agnostic composite offers a simple and replicable operationalization as a preclinical indicator of health.

At the same time, although the current findings identified relatively better and worse analytic strategies in terms of prediction of health outcomes, these patterns should not be taken as evidence about underlying mechanisms. Stronger or more reliable associations with health outcomes do not imply that the set of biomarkers is more causally involved in disease processes. Circulating biomarkers, particularly those influenced by non-immune sources (e.g., CRP), do not directly reflect inflammatory activity at the site of disease, and this interpretative challenge becomes more pronounced as larger numbers of biomarkers are combined as composites. Therefore, these findings may be best viewed as an impetus for future work examining analytical choices using experimental, in-vitro, or animal paradigms better suited for delineating causal mechanisms, rather than as a prescription for best practices in analyzing inflammatory biomarkers or as evidence for which biomarkers are most mechanistically relevant for health.

Future research may consider other common analytical practices (e.g., outlier handling, assaying platforms, multiplex panel composition), use more systematic approaches (e.g., multiverse or specification curve analyses; Rengasamy et al., 2023; Simonsohn et al., 2020), and expand evaluative criteria to include biological alignment (e.g., correspondence with established immune processes like LPS-stimulated cytokine production in PBMC cell cultures, rather than non-immune correlates like adiposity). Psychometric criteria (e.g., internal consistency) offer another perspective, but future research should clarify whether applying thresholds developed for psychological measures (e.g., survey items) is appropriate for biological data, and whether context-specific thresholds should be developed.

Moreover, future research may use cumulative data approaches to build larger samples and broader biomarker panels. Studies commonly report bivariate inter-biomarker correlation matrices, which can be *meta*-analytically synthesized to create a pooled matrix, spanning a larger set of biomarkers than any single study includes. This pooled

matrix provides more generalizable correlation estimates that can then be used as input for meta-analytic factor analysis (Cheung, 2015). Generalizability and replicability of factor structure can then be tested across coded sample or methodological characteristics. Alternatively, progress may come from pooling individual-level inflammatory biomarker data across studies, along with demographic and biological variables that would allow evaluation of generalizability and predictive validity. This integrative data approach would yield greater statistical power needed (Curran and Hussong, 2009) to conduct systematic analyses for evaluating multiple analytical decisions.

4.5. Limitations

First, our evaluation assumed nonzero inflammation-health associations; if true correlations were zero (as in some populations (McDade, 2023), the observed benefits of log-transformation and composites may instead reflect unaccounted confounding (e.g., here, the analytical differences reduced in exploratory analyses controlling for age, sex, and race). Moreover, biomarkers were also treated as indicators of inflammation, although some can be elevated independent of inflammatory activities (Del Giudice and Gangestad, 2018). Moreover, basal levels of some circulating biomarkers may be weak predictors in heterogeneous samples like MIDUS, with potentially different biomarker clustering than in acute disease states. Future research should test these findings using experimental paradigms (e.g., viral challenge) or clinical samples characterized by elevated inflammatory activity. Second, we used Bonferroni adjustments, which are more conservative. Future work should assess robustness to other corrections (e.g., FDR). Third, a limitation of using PLS-SEM to construct formative composites is that indicator weights were estimated to maximize predictions of the very health outcomes used to evaluate the composites. Sensitivity analyses excluding formative composites yielded similar conclusions, but future methodological work is needed to develop alternative approaches for constructing formative composites. Fourth, we only included eight biomarkers (most of which reflect innate immune pathways) and eight self-reported health outcomes, for which longitudinal new incidence was relatively low. Future research should include broader biomarker panels and objective or incident health outcomes. Finally, compared to U.S. census estimate at the time of data collection, the analytical samples underrepresented participants of color, male at birth, and younger adults, limiting generalizability. Future research should test measurement invariance and replicability across demographic and methodological characteristics.

CRediT authorship contribution statement

Phoebe H. Lam: Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Conceptualization. **Gregory E. Miller:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

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

Acknowledgements

The authors have no conflicts of interest or financial relationships relevant to this article to disclose. The authors would like to thank the Reviewers for their thoughtful feedback, which strengthened the manuscript's theoretical framing and interpretations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbi.2026.106806.

Data availability

The authors do not have permission to share data. Data is publicly available here: icpsr.umich.edu/web/ICPSR/series/203

References

Abbas, A.K., Lichtman, A.H., Pillai, S., 2014. *Cellular and Molecular Immunology E-Book: Cellular and Molecular Immunology E-Book*. (Elsevier Health Sciences, 2014).
Aguirre-Urreta, M.I., Marakas, G.M., 2008. The use of PLS when analyzing formative constructs: Theoretical analysis and results from simulations. https://aisel.aisnet.org/icis2008/185/.
Albrecht, U., et al., 2007. Activation of NF-κB by IL-1β blocks IL-6-induced sustained STAT3 activation and STAT3-dependent gene expression of the human γ-fibrinogen gene. Cell. Signal. 19, 1866–1878.
Bottigliengo, D., et al., 2022. A Mendelian randomization study investigating the causal role of inflammation on Parkinson's disease. Brain 145, 3444–3453.
Brim, O.G., *et al.*, 2020. Midlife in the United States (MIDUS 1), 1995-1996. Inter-university Consortium for Political and Social Research [distributor] https://doi.org/10.3886/ICPSR02760.v19.
Brown, T.A., 2015. *Confirmatory Factor Analysis for Applied Research, 2nd Ed.* xvii, 462 (The Guilford Press, New York, NY, US).
Chae, W.-J., et al., 2010. Ablation of IL-17A abrogates progression of spontaneous intestinal tumorigenesis. PNAS 107, 5540–5544.
Cheung, M.W.-L., 2015. Meta-Analysis: A Structural Equation Modeling Approach. John Wiley & Sons.
Chiang, J.J., Lam, P.H., Chen, E., Miller, G.E., 2022. Psychological stress during childhood and adolescence and its association with inflammation across the lifespan: a critical review and meta-analysis. Psychol. Bull. 148, 27.
Curran, P.J., Hussong, A.M., 2009. Integrative data analysis: the simultaneous analysis of multiple data sets. Psychol. Methods 14, 81.
Danesh, J., et al., 2000. Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. BMJ 321, 199–204.
Danesh, J., et al., 2008. Long-term interleukin-6 levels and subsequent risk of coronary heart disease: two new prospective studies and a systematic review. PLoS Med. 5, e78.
Del Giudice, M., Gangestad, S.W., 2018. Rethinking IL-6 and CRP: why they are more than inflammatory biomarkers, and why it matters. Brain Behav. Immun. 70, 61–75.
De Vries, J.J., Snoek, C.J.M., Rijken, D.C., De Maat, M.P.M., 2020. Effects of post-translational modifications of fibrinogen on clot formation, clot structure, and fibrinolysis: a systematic review. ATVB 40, 554–569.
Dinarello, C.A., 2011. Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. Blood 117, 3720–3732.
Edwards, J.R., Bagozzi, R.P., 2000. On the nature and direction of relationships between constructs and measures. Psychol. Methods 5, 155.
Egnot, N.S., et al., 2018. An exploratory factor analysis of inflammatory and coagulation markers associated with femoral artery atherosclerosis in the San Diego population study. Thromb. Res. 164, 9–14.
Fragão-Marques, M., Francisca-Marques, M., Rocha Neves, J., Ozben, T., 2025. Association of inflammatory biomarkers with morbidity and mortality risk in patients with peripheral artery disease: a systematic review and -meta-analysis. Crit. Rev. Clin. Lab. Sci. 1–20. https://doi.org/10.1080/10408363.2025.2512472.
Furman, D., et al., 2019. Chronic inflammation in the etiology of disease across the life span. Nat. Med. 25, 1822–1832.
García-Pérez, M.A., 2023. Use and misuse of corrections for multiple testing. Methods Psychol. 8, 100120.
Gasparoto, A., Martinez, T., Saldanha, A., Margeotto, A., 2021. Fibrinogen as a cardiovascular risk marker. J. Internal Med. Cardiovasc. Res. 1, 1–7.
Guo, Y.-Z., Pan, L., Du, C.-J., Ren, D.-Q., Xie, X.-M., 2013. Association between C-reactive protein and risk of cancer: a meta-analysis of prospective cohort studies. Asian Pac. J. Cancer Prev. 14, 243–248.
Haddad, J.J., 2002. Cytokines and related receptor-mediated signaling pathways. Biochem. Biophys. Res. Commun. 297, 700–713.
Haraldsen, G., Kvale, D., Lien, B., Farstad, I.N., Brandtzaeg, P., 1996. Cytokine-regulated expression of E-selectin, intercellular adhesion molecule-1 (ICAM-1), and vascular cell adhesion molecule-1 (VCAM-1) in human microvascular endothelial cells. J. Immunol. 156, 2558–2565.
Harris, T.B., et al., 1999. Associations of elevated interleukin-6 and C-reactive protein levels with mortality in the elderly. Am. J. Med. 106, 506–512.
Hoogland, I.C.M., Houbolt, C., Van Westerloo, D.J., Van Gool, W.A., Van De Beek, D., 2015. Systemic inflammation and microglial activation: systematic review of animal experiments. J. Neuroinflammation 12, 114.
Hu, L.T., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct. Equ. Model. Multidiscip. J. 6, 1–55.
Ioannidis, J.P.A., 2014. Discussion: why "an estimate of the science-wise false discovery rate and application to the top medical literature" is false. Biostatistics 15, 28–36.
Kaptoge, S., et al., 2010. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. Lancet 375, 132–140.

Khudiakova, A.D., et al., 2023. Associations of adipocytokines with the development of cardiovascular events in young people. J. Personalized Med. 13, 1582.
Koukkunen, H., et al., 2001. C-reactive protein, fibrinogen, interleukin-6 and tumour necrosis factor-α in the prognostic classification of unstable angina pectoris. Ann. Med. 33, 37–47.
Kristono, G.A., et al., 2020. An IL-6-IL-8 score derived from principal component analysis is predictive of adverse outcome in acute myocardial infarction. Cytokine X 2, 100037.
Lai, K., 2018. Estimating standardized SEM parameters given nonnormal data and incorrect model: methods and comparison. Struct. Equ. Model. Multidiscip. J. 25, 600–620.
Lai, K.S.P., et al., 2017. Peripheral inflammatory markers in Alzheimer's disease: a systematic review and meta-analysis of 175 studies. J. Neurol. Neurosurg. Psychiatry 88, 876–882.
Li, Y., et al., 2017. Hs-CRP and all-cause, cardiovascular, and cancer mortality risk: a meta-analysis. Atherosclerosis 259, 75–82.
MacCallum, R.C., Zhang, S., Preacher, K.J., Rucker, D.D., 2002. On the practice of dichotomization of quantitative variables. Psychol. Methods 7, 19.
Marcos-Pérez, D., et al., 2020. Association of inflammatory mediators with frailty status in older adults: results from a systematic review and meta-analysis. GeroScience 42, 1451–1473.
McDade, T.W., 2023. Three common assumptions about inflammation, aging, and health that are probably wrong. Proc. Natl. Acad. Sci. USA 120, e2317232120.
McNeill, E., Channon, K.M., Greaves, D.R., 2010. Inflammatory cell recruitment in cardiovascular disease: murine models and potential clinical applications. Clin. Sci. 118, 641–655.
Medzhitov, R., 2021. The spectrum of inflammatory responses. Science 374, 1070–1075.
Michels, N., van Aart, C., Morisse, J., Mullee, A., Huybrechts, I., 2021. Chronic inflammation towards cancer incidence: a systematic review and meta-analysis of epidemiological studies. Crit. Rev. Oncol. Hematol. 157, 103177.
Miller, G.E., 2009. In search of integrated specificity: Comment on Denson, Spanovic, and Miller (2009). https://psycnet.apa.org/record/2009-19763-002.
Misra, B.B., Langefeld, C., Olivier, M., Cox, L.A., 2019. Integrated omics: tools, advances and future approaches. J. Mol. Endocrinol. 62, R21–R45.
Moriarity, D.P., 2022. A primer on common analytic concerns in psychoneuroimmunology: Alternatives and paths forward. Brain Behav. Immun. 102, 338.
Moriarity, D.P., Horn, S.R., Kautz, M.M., Haslbeck, J.M., Alloy, L.B., 2021. How handling extreme C-reactive protein (CRP) values and regularization influences CRP and depression criteria associations in network analyses. Brain Behav. Immun. 91, 393–403.
Moriarity, D.P., Ellman, L.M., Coe, C.L., Olino, T.M., Alloy, L.B., 2021. A physiometric investigation of inflammatory composites: comparison of "a priori" aggregates, empirically-identified factors, and individual proteins. Brain Behav. Immunity-Health 18, 100391.
Nakano-Narusawa, Y., et al., 2021. Relationship between lung carcinogenesis and chronic inflammation in rodents. Cancers 13, 2910.
Nathan, C., Ding, A., 2010. Nonresolving inflammation. Cell 140, 871–882.
Ngwa, D.N., Pathak, A., Agrawal, A., 2022. IL-6 regulates induction of C-reactive protein gene expression by activating STAT3 isoforms. Mol. Immunol. 146, 50–56.
Pearson, T.A., et al., 2003. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. Circulation 107, 499–511.
Pepys, M.B., Hirschfield, G.M., 2003. C-reactive protein: a critical update. J. Clin. Invest. 111, 1805–1812.
Picard, C., et al., 2003. Pyogenic bacterial infections in humans with IRAK-4 deficiency. Science 299, 2076–2079.
Pradeu, T., Thomma, B.P., Girardin, S.E., Lemaitre, B., 2024. The conceptual foundations of innate immunity: taking stock 30 years later. Immunity 57, 613–631.
Pripp, A.H., Stanišić, M., 2014. The correlation between pro-and anti-inflammatory cytokines in chronic subdural hematoma patients assessed with factor analysis. PLoS One 9, e90149.
Qin, P., Ho, F.K., Celis-Morales, C.A., Pell, J.P., 2025. Association between systemic inflammation biomarkers and incident cardiovascular disease in 423,701 individuals: evidence from the UK biobank cohort. Cardiovasc. Diabetol. 24, 162.
Rengasamy, M., Moriarity, D., Kraynak, T., Tervo-Clemmens, B., Price, R., 2023. Exploring the multiverse: the impact of researchers' analytic decisions on relationships between depression and inflammatory markers. Neuropsychopharmacology 48, 1465–1474.
Rubin, M., 2024. Inconsistent multiple testing corrections: the fallacy of using family-based error rates to make inferences about individual hypotheses. Methods Psychol. 10, 100140.
Ryff, C.D., *et al.*, 2017. Midlife in the United States (MIDUS Refresher 1), 2011-2014. Inter-university Consortium for Political and Social Research [distributor] https://doi.org/10.3886/ICPSR36532.v3.
Sarstedt, M., Ringle, C.M., Hair, J.F., 2022. Partial Least Squares Structural Equation Modeling. in *Handbook of Market Research* (eds Homburg, C., Klarmann, M. & Vomberg, A.) 587–632 (Springer International Publishing, Cham). doi:10.1007/978-3-319-57413-4_15.
Schett, G., Elewaut, D., McInnes, I.B., Dayer, J.-M., Neurath, M.F., 2013. How cytokine networks fuel inflammation: toward a cytokine-based disease taxonomy. Nat. Med. 19, 822–824.
Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis. Nat. Hum. Behav. 4, 1208–1214.
Tziakas, D.N., et al., 2007. Inflammatory and anti-inflammatory variable clusters and risk prediction in acute coronary syndrome patients: a factor analysis approach. Atherosclerosis 193, 196–203.
Unger, J.M., et al., 2025. When to adjust for multiplicity in cancer clinical trials. JNCI Monogr. 2025, 3–9.
Vandebergh, M., Becelaere, S., Group, C.I.W., Dubois, B., Goris, A., 2022. Body mass index, interleukin-6 signaling and multiple sclerosis: a mendelian randomization study. Front. Immunol. 13, 834644.
Vestweber, D., 2012. Relevance of endothelial junctions in leukocyte extravasation and vascular permeability. Ann. N. Y. Acad. Sci. 1257, 184–192.
Wei, T., et al., 2023. Circulating levels of cytokines and risk of cardiovascular disease: a Mendelian randomization study. Front. Immunol. 14, 1175421.
Weinstein, M., Ryff, C.D., Seeman, T.E., 2019. Midlife in the United States (MIDUS Refresher 1): Biomarker Project, 2012-2016. Inter-university Consortium for Political and Social Research [distributor] https://doi.org/10.3886/ICPSR36901.v6.
Zhu, M., et al., 2022. C-reactive protein and cancer risk: a pan-cancer study of prospective cohort and Mendelian randomization analysis. BMC Med. 20, 301.