A two-week plant-based challenge reduced animal protein consumption by roughly a third compared to a control group: a randomized evaluation with daily food logs

Mark Xu Sidney Hough Charles Lien Akash Borde Grant Hough

July 2026

Abstract

Plant-based "challenge" programs, short commitments to reduce or eliminate animal products, are among the most widely used tools for shifting diets, yet their evidence base consists almost entirely of uncontrolled surveys of self-selected participants. We report a randomized evaluation of a two-week challenge: of 541 sign-ups, the 458 who completed the initial survey were assigned to attempt the challenge or to eat as usual, and both arms logged the animal products they ate every day, which a validated language-model pipeline converted to grams of animal protein. Among the 274 participants who logged consistently enough for analysis, the treatment group logged 10.6 fewer grams per day than control after adjusting for baseline (95% CI $[4.3, 17.0]$, $p = .001$), a reduction of 31%. Despite being asked to eat normally, the median control participant ate 18% below their self-reported baseline. Participants found cravings and social meals harder than the protein and nutrition worries they had anticipated. Thirty days after the study ended, 83% of the 274 responded to a follow-up survey: the treatment arm remained 9.1 g/day below control, but the difference was no longer significant (95% CI $[-23.8, 5.5]$, $p = .22$).

1 Introduction

The consumption of meat and animal products has large consequences for public health, the climate, and animals. Processed meat is classified as carcinogenic and red meat as probably carcinogenic (Bouvard et al., 2015), and diets heavy in red and processed meat carry elevated risks of cardiovascular disease and type 2 diabetes (GBD 2017 Diet Collaborators, 2019). Animal agriculture generates a disproportionate share of food-system greenhouse-gas emissions and occupies most of the world's farmland (Poore and Nemecek, 2018). More than 80 billion land animals are killed for food each year, most of them raised industrially (Ritchie et al., 2023; Bonnet et al., 2020).

Organizations working to reduce animal-product consumption widely use the *challenge*: a time-boxed commitment to reduce or eliminate animal products, supported by a program's encouragement and resources (Grassian, 2020). Veganuary alone reports millions of cumulative sign-ups (Veganuary, 2025); Challenge 22 (Animals Now, 2026) and No Meat May (No Meat May, 2026) run continuously. Yet the evidence that challenges change what participants eat has two specific weaknesses. First, existing evaluations are almost entirely uncontrolled pre/post surveys of self-selected participants (Grassian, 2020; Faunalytics, 2020; Becker and Lawrence, 2022): without a comparison group, they cannot separate the program's effect from what motivated volunteers would have done anyway. Second, outcomes are typically retrospective self-reports, vulnerable to recall error and social-desirability bias; Mathur et al. (2021) estimated that such bias alone can produce apparent effects comparable in size to those reported in this literature.

To our knowledge, only one randomized trial of a structured challenge has been published. Marty et al. (2025) randomized 366 French university students who signed up for a one-month meat-reduction challenge to start immediately ($n = 187$) or after a waitlist ($n = 179$), and found meat consumption 34 g/day lower in the challenge arm during the challenge month (measured by food-frequency questionnaire); the between-arm difference was no longer significant three months later, though the challenge arm remained about 50 g/day below its own baseline. The nearest analogues outside the challenge frame are multicomponent behavioral programs, which combine self-monitoring, goal-setting, feedback, or meat substitutes and have produced comparable reductions in randomized evaluations (Frie et al., 2022; Bianchi et al., 2022). Across the wider meat-reduction literature, measured-consumption outcomes remain rare relative to attitudes and intentions (Green et al., 2025).

We report a randomized evaluation of a two-week plant-based challenge with 274 analyzed participants. The primary outcome is daily logged animal protein in grams, counting all animal products (meat, seafood, dairy, and eggs). Every day for two weeks, participants described what they had eaten in free-form text messages or website entries; a language-model pipeline parsed each log and a fixed conversion table mapped it to grams (Section 2.4). Day-resolved intake measurement has historically required burdensome diaries, structured recalls, or app-bound logging (Tansey et al., 2016), or has substituted purchase records for consumption (Aiello et al., 2020). Recently, the use of language models to code free-text food logs has been validated against expert coders (Zhou et al., 2026; Ase et al., 2026); to our knowledge, this study is the first to use such logs to measure the outcome of a randomized trial rather than to benchmark the coding itself. Control participants also submitted daily logs while eating normally, which lets us estimate the bias in self-reported baselines that single-arm evaluations absorb into their effect estimates (Section 3.4). We report the effect during the challenge, what participants found hard, who reduced most, and follow-through at 30 days (Section 3.7).

2 Methods

2.1 Recruitment and participants

Participants were recruited in May and June 2026 through community outreach (direct invitations, social media, and flyers) and distribution partners (EarthDay.org, NutritionFacts.org, Switch4Good, Zero Waste Chef, and Sabrina Pare), and enrolled at plantbasedstudy.org. Participation was unpaid and open to adults willing to attempt a two-week reduction; people already vegan, under 18, unable to respond to daily messages in English, or for whom participation was medically inadvisable (including current eating disorders and pregnancy without physician clearance) were excluded. An initial survey covered age, gender, country, dietary pattern (omnivore, pescatarian, or vegetarian), and motivation; elicited usual consumption in servings per typical day of meat ("about the size of a deck of cards"), dairy ("1 cup of milk or 2 slices of cheese"), and eggs; and asked participants to rate the expected difficulty of reducing (1 = not at all, 5 = extremely difficult) and to describe expected difficulties in free text. A parallel survey at study end repeated the difficulty questions for the difficulties participants actually experienced. Thirty days after the study ended, a further survey repeated the enrolment consumption questions and asked participants to assess how their consumption had changed since before the study.

2.2 Study design and randomization

Of the 541 people who signed up, the 458 who completed the initial survey were randomized within six strata: gender crossed with three consumption groups (omnivores above the omnivore median of baseline meat consumption, omnivores below it, and non-omnivores). Within each stratum, 80% were assigned to treatment and 20% to control ($n = 365$ and 93). The treatment-heavy allocation reflects the study's secondary interest in within-treatment questions; the control arm was sized to provide a reliable between-arm benchmark.

2.3 Intervention

Treatment participants were told to reduce their animal-product consumption as much as they could for 14 days. They received a short text guide that emphasized the goal was to reduce consumption of animal products, not eliminate it, and provided practical advice for how to do so. Control participants were told to eat normally and did not receive a guide. Both arms received the same daily logging prompts (two texts per day, or one email for web loggers; Section 2.4), framed as a "plant-based trial" for treatment and a "food diary" for control.

2.4 Measurement: daily food logs

Participants who enrolled with a phone number received two daily texts: a morning prompt to reply through the day with the animal products (meat, dairy, eggs) they ate, and an evening prompt to log anything missed. Those who enrolled by email received one daily email and logged on the study website, where anyone could fill in missed days. Responses were converted to grams by a two-stage language-model pipeline: a language model (Claude Sonnet 4.6) extracted each day's animal-product items (food, quantity, unit); a fixed, hand-curated table then mapped each item to a low–high range of standard servings and to grams of animal protein, at 26 g per serving of meat (85 g cooked), 20 g per serving of seafood, 8.5 g per serving of dairy (1 cup of milk or equivalent), and 6.3 g per egg. A participant's daily total is the midpoint of the range summed across categories; endpoint consumption is the mean of their usable days. Baseline consumption in grams was computed from the initial survey's servings-per-day questions, applying the same per-serving protein values.[1]

2.5 Validation of the language-model pipeline

The first author blind-coded 200 randomly sampled daily messages (163 treatment, 37 control) into servings per category, without access to the pipeline's codes. On whether a category was consumed at all, human–pipeline agreement was 96.5–99% (Cohen's $\kappa = 0.92$–0.96 across meat, seafood, dairy, and eggs), with agreement on 26 of the 27 days the pipeline scored as zero. On quantities, daily grams correlated at $r = 0.88$ (ICC = 0.87; mean absolute difference 6.7 g). The

[1] The initial survey did not ask about seafood separately; participants plausibly folded fish into "meat." Baseline enters the analysis only as a covariate, identically for both arms, so the asymmetry does not affect the between-arm contrast.

pipeline ran 1.8 g/day above the human coder, likely from taking range midpoints for vaguely quantified foods. The between-arm difference in this error was 2.4 g/day (pipeline higher on control logs; 95% CI $[-2.1, +6.8]$, Welch $p = .31$, $n = 37$ control messages): not distinguishable from zero, and even at face value accounting for roughly a fifth of the treatment effect, leaving about 8 g/day. These agreement levels are comparable to recent validations of language-model food coding (90–94% expert agreement and micro-F1 of 0.82; Ase et al., 2026; Zhou et al., 2026).

2.6 Analysis sample

Participants who never submitted a daily log, or who had three or more non-response days out of 14, were excluded from the consumption analyses under the study's data-quality rule. The primary analysis is therefore per-protocol (an effect among participants who kept logging) rather than intention-to-treat; Section 2.7 describes the imputation analysis we use to bound the influence of attrition, and Section 3.1 reports attrition by arm. Figure 1 traces the flow from sign-up to analysis in both arms.

2.7 Statistical analysis

The primary outcome was endpoint consumption: a participant's mean daily animal protein in grams across their usable study days. We estimated the per-protocol effect of assignment to the challenge using ordinary least squares, regressing endpoint consumption on treatment assignment with adjustment for baseline consumption (ANCOVA) and heteroskedasticity-consistent robust standard errors. We report the treatment coefficient with its 95% confidence interval and p-value. For binary outcomes (for example, whether a participant reduced at all), we report the share of participants with the outcome in each arm and the difference between them. Because baseline enters as a covariate, adjusted models use the 273 analyzed participants with a reported baseline.

We assessed the robustness of the primary result in four ways. First, because dietary outcomes are typically right-skewed, we tested the unadjusted between-arm difference with a Mann–Whitney rank test, which is insensitive to outliers. Second, to bound the influence of attrition, we re-ran the primary model assuming that every excluded participant with a reported baseline consumed exactly at that baseline ($n = 415$ of 458; the remaining 43 lacked a reported baseline). This no-change imputation attenuates the effect when applied to excluded treatment participants and inflates it when applied to excluded controls; because exclusions ran roughly four-to-one treatment, it is conservative on net. Third, because the three-non-response-day rule is an operational data-quality screen rather than a substantive choice, we re-estimated the primary model at every alternative threshold, from admitting only participants with no non-response days to admitting anyone with at least one usable day. Fourth, we reprocessed the free-form logs into grams using a pipeline variant that estimated food-specific protein values rather than applying the fixed per-category values; this changed totals and the treatment effect by less than 10% and altered no conclusion. Secondary and subgroup analyses were not corrected for multiplicity and should be read descriptively.

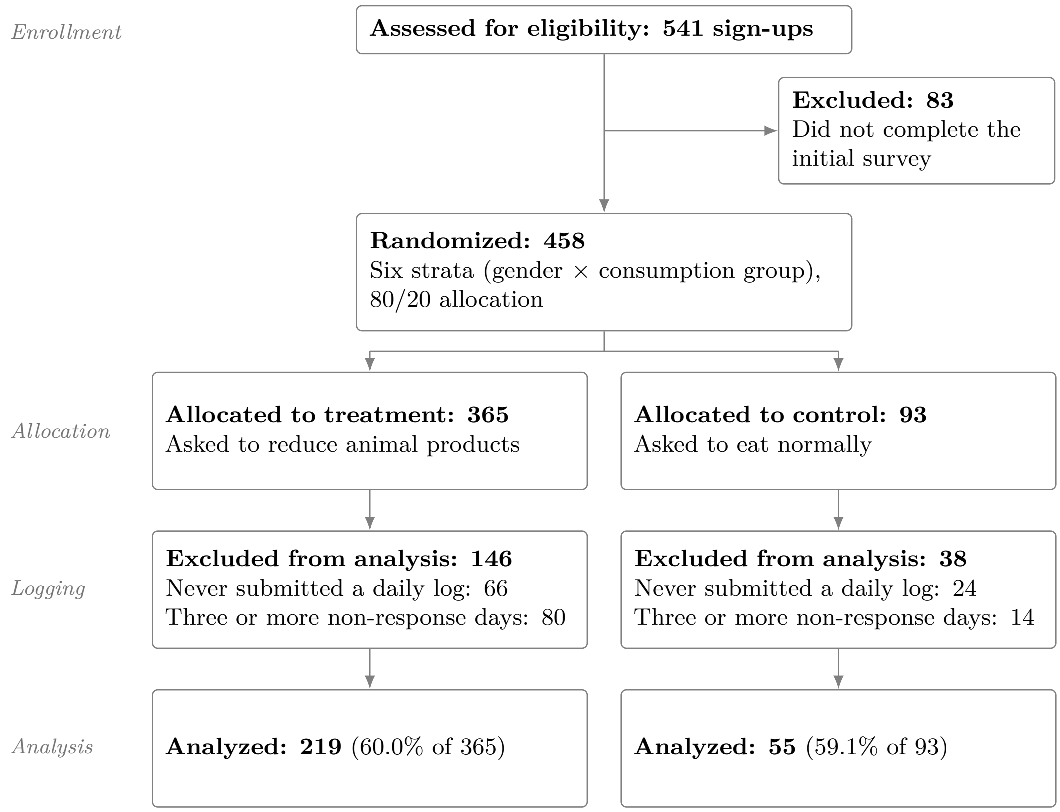

Figure 1: Participant flow from sign-up to analysis. Because allocation was 80/20, the treatment arm is roughly four times the size of the control arm and loses more participants in absolute terms at every stage, while the share retained is nearly identical (60.0% vs. 59.1%, χ^2 $p = .97$). A non-response day is a study day with no log or an unintelligible one; participants who never submitted a log of any kind are counted separately from those excluded by the three-non-response-day rule. Baseline-adjusted models use the 273 analyzed participants with a reported baseline (219 treatment, 54 control).

3 Results

3.1 Participant flow and characteristics

Figure 1 traces participant flow. Of the 458 randomized participants (365 treatment and 93 control, reflecting the four-to-one allocation of Section 2.2), 90 never submitted a daily log (66 treatment, 24 control) and 94 more (80 treatment, 14 control) were excluded under the data-quality rule, leaving 274 for analysis (219 treatment, 55 control). The treatment arm lost more participants in absolute terms simply because it was four times larger: inclusion rates were nearly identical across arms, 60.0% of treatment (219 of 365) and 59.1% of control (55 of 93; χ^2 $p = .97$), and excluded participants did not detectably differ from included ones in baseline consumption ($p \geq .48$ in each arm). Equal rates do not by themselves rule out differential selection, since the two arms may stop logging for different reasons and consumption among those who stopped is unobserved; the imputation and threshold analyses in Section 3.2 bound the observable channels. Randomization held in the analyzed sample: baseline animal protein was balanced (46.6 vs. 46.3 g/day, $p = .96$), as were age, gender, and dietary pattern.

The analyzed sample skews female (73%) and middle-aged (mean age 51, SD 18), with about 80% residing in the United States; 73% were omnivores, 18% vegetarians, and 9% pescatarians. The most common motivations were environmental impact (34%), health (27%), curiosity (19%), and animal welfare (15%). This is an older, more heterogeneous population than the student samples common in this literature (e.g., Marty et al., 2025; Jalil et al., 2023; Haile et al., 2021).

3.2 Effect of the challenge on animal protein consumption

Treatment participants logged a mean of 22.6 g of animal protein per day (median 19.1) against the control group's 33.8 (median 30.4). Adjusting for baseline, the challenge reduced consumption by 10.6 g/day (95% CI $[4.3, 17.0]$, $p = .001$): 31% of the control group's intake, or in everyday units about two-fifths of a serving of meat per day, roughly three meat-serving-equivalents per week. The unadjusted difference is nearly identical (11.2 g/day; Hedges' $g = 0.56$ favoring treatment) and holds under the rank-based test (Mann–Whitney $p < .001$). The effect survives the no-change imputation (Section 2.7), at 6.7 g/day (95% CI $[0.7, 12.8]$), and does not depend on the exclusion threshold: re-estimating the model at every threshold from admitting only participants with no non-response days ($n = 237$) to admitting anyone with at least one usable day ($n = 361$) yields effects between 10.4 and 11.9 g/day, all $p < .002$.

Relative to their own starting points (Figure 2), the median treatment participant ate 47% of their self-reported baseline, versus 82% for the median control.[2] Treatment participants were more likely to reduce at all (84% vs. 63%; χ^2 $p = .001$), and descriptively more likely to cut over 10 g/day (66% vs. 41%) and to end under 10 g/day (28% vs. 13%). The reduction came mostly from meat (6.5 of the 11.2 g/day unadjusted difference), then dairy (2.6 g), seafood (1.4 g), and eggs (0.7 g, not statistically significant).

3.3 Consumption trajectories over the two weeks

Daily consumption by arm was stable across the study (Figure 3): treatment averaged 23.0 g/day in week one and 22.2 in week two (paired $p = .34$); controls 33.1 and 34.4 ($p = .49$). Day-level missingness was low and similar across arms (1.3% of treatment person-days, 0.9% of control), so selective skipping of high-consumption days is unlikely to drive the pattern.

3.4 The control arm's drop from baseline: a benchmark for single-arm designs

Control participants, despite being asked to eat normally, consumed below their self-reported baseline: the median ate 82% of it (Wilcoxon signed-rank $p = .014$, $n = 54$ controls with a reported baseline).[3] Four mechanisms plausibly contribute: control participants deliberately reducing animal-product consumption despite being told to eat as normal; overestimation of usual consumption in a one-shot survey; the reducing effect of daily logging itself (self-monitoring is an active ingredient in dietary interventions; Frie et al., 2022); and different calibration of the baseline and endpoint instruments. The two-week design cannot separate

[2] Percent-of-baseline statistics use the sample in Figure 2's caption; they are descriptive, with the causal claim resting on the between-arm contrast.

[3] Because ratios and differences have different distributions, the median *drop* (4.6 g/day) is smaller than 18% of the median baseline; the within-person mean drop (13.2 g/day, $n = 54$) is inflated by a skewed tail.

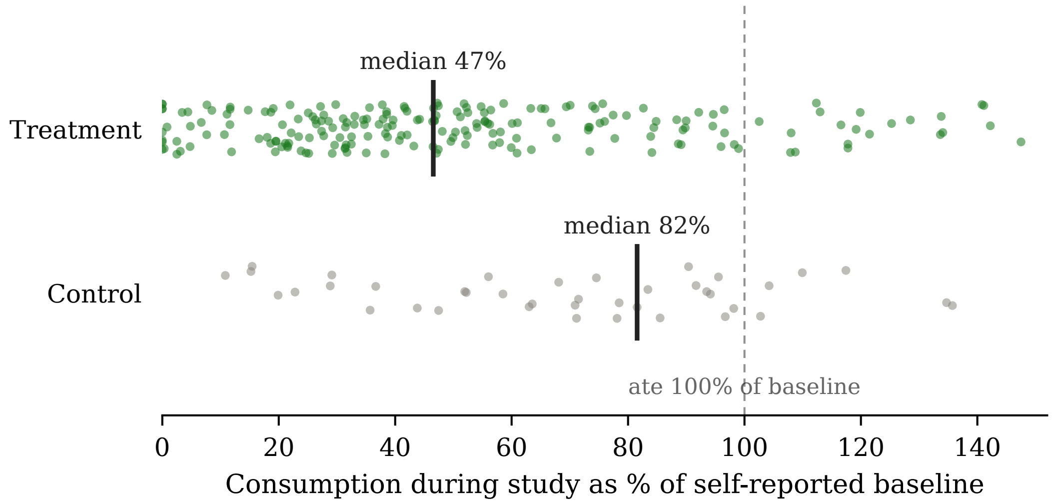

Figure 2: Consumption during the study as a share of each participant's self-reported baseline; dots are participants, vertical bars group medians. Medians use the 259 analyzed participants with baselines of at least 5 g/day (210 treatment, 49 control; 14 below threshold and 1 with no baseline excluded). The 15 participants above 150% (6 treatment, 9 control) are hidden from the display but included in the medians.

these mechanisms, but it measures their combined size: roughly 18% of baseline at the median. The 30-day follow-up (Section 3.7) separates them partially, and moves weight away from the first. Within the control arm, percent-of-baseline correlates -0.43 with baseline itself, the participants who reported the most at enrollment showing the largest apparent drops—the signature of a noisy one-shot recall rather than of behavior. And once control participants are measured again on the instrument that collected their baseline, they return to within a few percent of it: the same controls average 51.3 g/day at enrollment, 36.1 while logging daily, and 48.3 on the 30-day survey.

Uncontrolled evaluations that compare participants to their own recalled baseline risk overestimating their headline effects: the treatment arm's within-person drop (median 20 g/day) is roughly twice the between-arm causal estimate, so a single-arm before–after reading of our own treatment arm would have doubled the headline effect.

3.5 Expected and experienced difficulties

At enrollment, treatment participants rated how difficult they expected reducing to be and described the difficulties they anticipated in free text; at study end, they rated and described the difficulties they had actually experienced. Free-text responses were coded into eleven difficulty themes (those shown in Figure 4), with multiple themes allowed per response and the same codebook applied to the before and after responses. Overall difficulty expectations were well calibrated: of the 202 treatment completers who rated both, 46% gave the same rating twice, 91% moved at most one point, and the median was 2 ("slightly difficult") both times. The *kind* of difficulty shifted (Figure 4). Participants named more difficulties after than before (2.0 vs. 1.4 mentions per respondent): cravings and taste rose from 27% to 44% of participants, eating out and social meals from 17% to 32%, finding good options from 21%

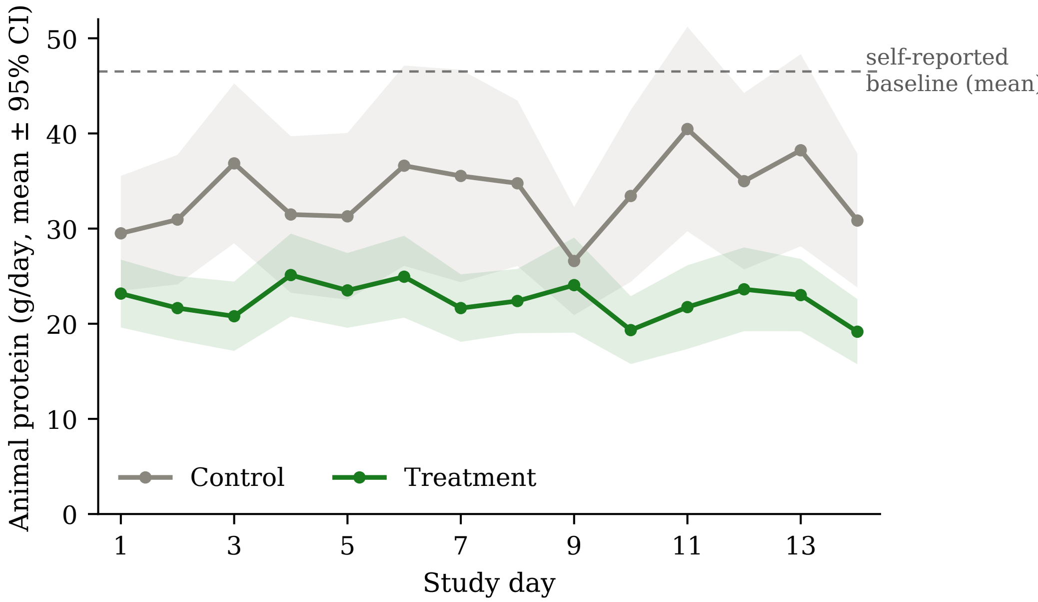

Figure 3: Mean daily animal protein by arm across the 14 study days, with 95% confidence bands. The dashed line marks the analyzed sample's mean self-reported baseline (46.5 g/day). Treatment $n = 219$, control $n = 55$.

to 31%. Protein and nutrition, among the most common worries beforehand (24%), rose only to 30%.

3.6 Predictors of reduction

Reduction size was similar across treatment-arm subgroups (same sample as Figure 2), with no significant differences. Men and women ended at a median 47% and 44% of baseline (Mann–Whitney $p = .84$); age groups 18–35, 36–55, and 56+ at 39%, 48%, and 51% (Kruskal–Wallis $p = .20$). More unexpectedly, difficulty did not track reduction either: expected difficulty (rated at enrollment), experienced difficulty (rated at study end), and the gap between the two were all essentially unrelated to how much a participant cut (Spearman $\rho = 0.05$, $p = .47$, $n = 210$ for expected; $\rho = 0.09$, $p = .21$, $n = 193$ for experienced; $\rho = 0.06$, $p = .42$, $n = 193$ for the gap). We had expected participants who found reducing harder to reduce less; instead, self-rated difficulty appears unrelated to reduction.

The treatment–control contrast was clearest among omnivores (unadjusted 12.6 g/day, $p < .001$) and, mechanically, small among the already low-consuming vegetarians (1.8 g/day). Subgroup analyses are underpowered, particularly in the small control arm, but nothing suggests that different demographic groups systematically reduce by different amounts.

3.7 Intentions and follow-through at 30 days

At study end, 85% of final-survey respondents ($n = 281$, including participants excluded from the consumption analyses) said they will probably (33%) or definitely (52%) incorporate more plant-based foods into their long-term diet (Figure 5). Thirty days after the study ended, all

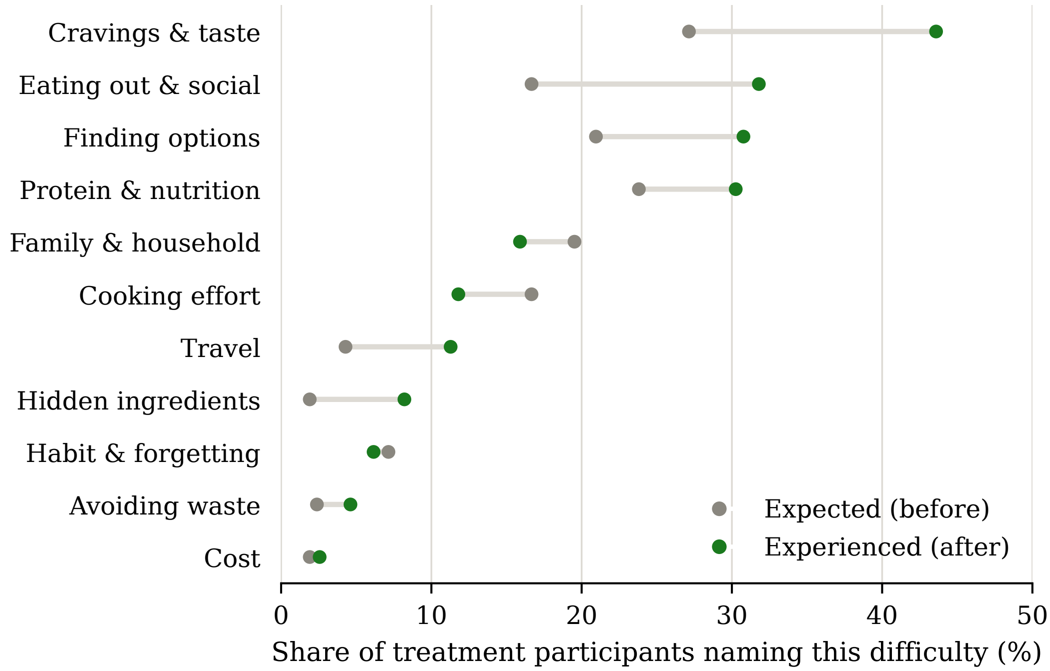

Figure 4: Share of treatment participants naming each difficulty before the study (expected) and after (experienced). Free-text responses, thematically coded; participants could name several. $n = 210$ before, 195 after.

participants received a follow-up survey mirroring the baseline consumption questions plus a self-assessment of change. Two hundred forty-seven of the 458 randomized participants (54%) responded, including 83% of the treatment and 86% of the control participants in the analyzed sample. Response rates were similar across arms, and responders did not differ from non-responders on either baseline or during-study consumption in either arm (all Mann–Whitney $p > .24$). Because the follow-up repeats the enrollment survey's servings-per-day questions, baseline and 30-day intake are measured on the same instrument, which the during-study logs are not.

Adjusting for baseline, the treatment arm reported 9.1 g/day less animal protein than control at 30 days, a difference that was not significant (95% CI $[-23.8, 5.5]$, $p = .22$, $n = 246$). On the log scale, which fits the skewed ratio better, the estimate is 0.77 times control (95% CI $[0.57, 1.03]$, $p = .078$; rank-based Hodges–Lehmann 0.81, 95% CI $[0.64, 1.01]$). The point estimates retain roughly half the during-challenge effect—between 41% and 63% depending on the scale—but the attenuation is significant only on the log scale ($p = .04$ against $p = .49$ in grams), so how much of the effect decayed is not resolved.

Within-arm changes look larger: the median treatment participant reported 76% of their own baseline (mean change -10.0 g/day, Wilcoxon $p < 10^{-8}$, $n = 196$) and the median control participant 80% (-1.6 g/day, $p = .13$, $n = 50$). Neither is an effect estimate. Baseline here is a single figure recalled at sign-up, and error in that figure is enough on its own to make whoever reported the most at sign-up appear to cut the most: in the control arm, the half with the lowest baselines reports 110% of theirs and the half with the highest reports 71% (Spearman

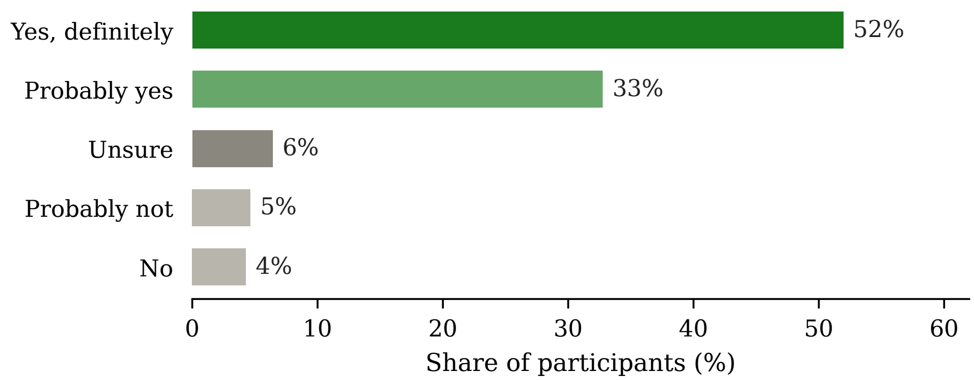

Figure 5: "Do you plan to incorporate more plant-based foods into your long-term diet as a result of the study?" All assigned participants completing the final survey, $n = 281$.

-0.43, $p = .002$, $n = 48$).

The self-assessment item separates the arms more cleanly than the consumption numbers do. 53% of treatment and 33% of control respondents said they now eat less animal protein than before the study (difference 20 percentage points, 95% CI $[5, 34]$, Fisher $p = .018$). This is a secondary, self-reported outcome, and we do not read it as evidence for a consumption difference the measured data do not show; it is, though, consistent in direction with every estimate above, and self-assessment tracked the numbers within respondents—those who said they ate less reported 17.6 g/day below baseline, against +0.7 g/day for everyone else. Stated intentions thus predicted the direction of behavior without predicting its magnitude (Sheeran and Webb, 2016).

The follow-up further demonstrates the importance of a comparison group. Taken as a single-arm before–after study—the design of nearly every published challenge evaluation—it would report that participants sustained a 24% reduction 30 days after a two-week program. The control arm, told to eat normally, shows a 20% apparent reduction over the same window on the same instrument.

4 Discussion

4.1 Principal findings

A two-week plant-based challenge, evaluated with random assignment and daily logged consumption, reduced logged animal protein intake by about a third relative to control. Thirty days after the challenge ended the two arms were no longer statistically distinguishable, with point estimates retaining roughly half the effect (Section 3.7).

In food-weight terms, 10.6 g/day of animal protein is the protein in roughly 35 g of cooked meat: about an ounce and a quarter per day, or some 17 ounces—close to six meat servings—across the two weeks. Because the outcome pools meat, seafood, dairy, and eggs, that is a protein-equivalent figure rather than a literal count of meat servings forgone; the meat-only share of the unadjusted difference (6.5 of 11.2 g/day) maps to about 21 g of cooked meat per day.

The effect is close to that of the one comparable trial. Marty et al. (2025) found a between-arm reduction of about 34 g/day of meat by food weight, roughly 10 g/day of animal protein at our conversion factor (26 g per 85 g cooked serving), against our 10.6 g/day; their estimate covers meat only, ours all animal products (11.2 g/day unadjusted, of which 6.5 from meat). Reproducing an effect of this size under stricter measurement (daily logs rather than a retrospective food-frequency questionnaire, a validated coding pipeline, and a control arm that logged identically) makes this study a more robust validation of the during-challenge effect than has previously existed.

Three features of the design support the causal reading. Random assignment removes self-selection: treatment and control participants were drawn from the same pool of sign-ups, so the difference between arms reflects the challenge rather than the kind of person who joins. The validation found very small differential coding error between arms, though it is estimated from only 37 control messages (Section 2.5). And attrition, though substantial, was nearly identical in rate across arms, unrelated to baseline consumption, and the effect survives both the no-change imputation and every alternative exclusion threshold (Section 3.2).

4.2 Implications for programs

For organizations running challenges, the basic result is that the core mechanism works on the population these programs reach: people who sign up and are asked to reduce do reduce, by large margins, for as long as the challenge runs. Program materials may over-weight protein reassurance relative to cravings and eating out, though two weeks cannot surface adequacy concerns that develop over months. We found no evidence that the effect is confined to particular demographics.

We also recommend that challenge programs build measurement into the program to reduce reliance on exit surveys. Daily free-text logging is light for participants, and automated coding makes analysis relatively cheap. Challenge programs might also consider randomizing a small fraction of sign-ups into a control arm to allow for simple analysis of the efficacy of the challenge.

4.3 Limitations

The chief limitation is that consumption is self-reported, and the two arms faced different reporting pressures. Treatment participants had committed to reducing, and for them a daily prompt about animal products could be interpreted as a compliance check; if some omitted foods to appear compliant, our estimate overstates the true effect. The validation bounds coding error, which was small and balanced across arms, but it cannot verify that participants reported all animal products they consumed.

Attrition is the second concern. About 40% of randomized participants logged too rarely to analyze, so the primary estimate is per-protocol. We think differential attrition is unlikely to account for a substantive portion of the effect: inclusion rates were nearly identical across arms (60.0% of the 365 assigned to treatment vs. 59.1% of the 93 assigned to control), excluded participants matched included ones on baseline consumption, and the effect survives both the imputation that assumes every excluded participant changed nothing (6.7 g/day) and every alternative log-completeness threshold (Section 3.2).

Several further limitations qualify the interpretation. The outcome metric, grams of animal protein, weights a cup of milk far below a serving of beef and cannot see substitutions within animal products (beef to chicken, meat to dairy) whose health, climate, and welfare consequences differ. Two weeks of measurement is also silent on persistence, the margin on which prior challenge effects have decayed, and the 30-day follow-up is weaker evidence than the trial itself: it is a single retrospective recall rather than a period of daily logging, its control arm contains only 48 respondents with a usable baseline, and the resulting interval is wide enough to contain both no effect and the full during-challenge effect. Finally, participants were volunteers recruited through plant-based, sustainability, and nutrition communities, mostly middle-aged American women. The effect may differ in other populations.

Author contributions

Mark Xu: methodology, validation, formal analysis, visualization, writing. **Sidney Hough:** conceptualization, methodology, resources, visualization. **Charles Lien:** software, data curation. **Akash Borde:** resources. **Grant Hough:** resources.

Acknowledgments

We thank Seth Ariel Green for advice on framing and analysis, and Chris Bryant of Bryant Research for additional advice. We thank our distribution partners: EarthDay.org, NutritionFacts.org, Switch4Good, Zero Waste Chef, and Sabrina Pare.

Data availability

De-identified participant-level data are available at `https://github.com/worldalliance/plant-based-study-data`. The raw free-text food logs are not released; the published data contain the coded gram totals derived from them.

References

Luca Maria Aiello, Daniele Quercia, Rossano Schifanella, and Lucia Del Prete. Tesco Grocery 1.0, a large-scale dataset of grocery purchases in London. *Scientific Data*, 7:57, 2020.

Animals Now. Challenge 22. `https://challenge22.com`, 2026. Accessed July 2026.

Aia Ase, Jacek Borowicz, Kamil Rakocy, and Barbara Piekarska. Large language models for real-world nutrition assessment: Structured prompts, multi-model validation and expert oversight. *Nutrients*, 18(1):23, 2026.

Elisa Becker and Natalia S. Lawrence. The relationship between meat disgust and meat avoidance—a chicken-and-egg problem. *Frontiers in Nutrition*, 9:958248, 2022.

Filippo Bianchi, Cristina Stewart, Nerys M. Astbury, Brian Cook, Paul Aveyard, and Susan A. Jebb. Replacing meat with alternative plant-based products (RE-MAP): a randomized controlled trial of a multicomponent behavioral intervention to reduce meat consumption. *The American Journal of Clinical Nutrition*, 115(5):1357–1366, 2022.

Céline Bonnet, Zohra Bouamra-Mechemache, Vincent Réquillart, and Nicolas Treich. Viewpoint: Regulating meat consumption to improve health, the environment and animal welfare. *Food Policy*, 97:101847, 2020.

Véronique Bouvard, Dana Loomis, Kathryn Z. Guyton, et al. Carcinogenicity of consumption of red and processed meat. *The Lancet Oncology*, 16(16):1599–1600, 2015.

Faunalytics. Challenge 22+ & veg*n recidivism: A Faunalytics case study. `https://faunalytics.org/challenge-22-vegn-recidivism-a-faunalytics-case-study/`, 2020.

Kerstin Frie, Cristina Stewart, Carmen Piernas, Brian Cook, and Susan A. Jebb. Effectiveness of an online programme to tackle individual's meat intake through self-regulation (OPTIMISE): A randomised controlled trial. *European Journal of Nutrition*, 61:1313–1322, 2022.

GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 393(10184): 1958–1972, 2019.

Daniel T. Grassian. The dietary behaviors of participants in UK-based meat reduction and vegan campaigns: A longitudinal, mixed-methods study. *Appetite*, 154:104788, 2020.

Seth Ariel Green, Benny Smith, and Maya B. Mathur. Meaningfully reducing consumption of meat and animal products is an unsolved problem: A meta-analysis. *Appetite*, 216:108233, 2025.

Menbere Haile, Andrew Jalil, Joshua Tasoff, and Arturo Vargas Bustamante. Changing hearts and plates: The effect of animal-advocacy pamphlets on meat consumption. *Frontiers in Psychology*, 12:668674, 2021.

Andrew J. Jalil, Joshua Tasoff, and Arturo Vargas Bustamante. Low-cost climate-change informational intervention reduces meat consumption among students for three years. *Nature Food*, 4:218–222, 2023.

Lucile Marty, Maëlys Biehlmann, Anaïs Louveau, et al. The "Eat Less Meat" one-month challenge: a randomized controlled trial of a meat reduction pledge intervention among French university students. *International Journal of Behavioral Nutrition and Physical Activity*, 22, 2025.

Maya B. Mathur, Jacob R. Peacock, Thomas N. Robinson, and Christopher D. Gardner. Effectiveness of a theory-informed documentary to reduce consumption of meat and animal products: Three randomized controlled experiments. *Nutrients*, 13(12):4555, 2021.

No Meat May. No Meat May. `https://www.nomeatmay.net`, 2026. Accessed July 2026.

Joseph Poore and Thomas Nemecek. Reducing food's environmental impacts through producers and consumers. *Science*, 360(6392):987–992, 2018.

Hannah Ritchie, Pablo Rosado, and Max Roser. Animal welfare. Our World in Data, `https://ourworldindata.org/animal-welfare`, 2023.

Paschal Sheeran and Thomas L. Webb. The intention–behavior gap. *Social and Personality Psychology Compass*, 10(9):503–518, 2016.

Wesley Tansey, Edward W. Lowe, and James G. Scott. Diet2Vec: Multi-scale analysis of massive dietary data. arXiv:1612.00388, 2016.

Veganuary. Veganuary campaign in review. `https://veganuary.com`, 2025. Accessed July 2026.

H. Zhou, L. Chow, L. Harnack, S. Panda, E. N. C. Manoogian, M. Li, Y. Xiao, and R. Zhang. NutriRAG: unleashing the power of large language models for food identification and classification through retrieval methods. *Journal of the American Medical Informatics Association*, 33(4):802–811, 2026.