STUDENT REPORT APPENDIX

One-day offsite retreats, productivity and morale

This document records the protocol approved by the submitter before synthetic dataset generation. It may be attached to the submitted report.

Generation record

PASS ID
MSIN0093-15716F

PASS STATUS
Consumed - expired for further dataset generation

DATASET RECEIPT
EF-B386E557

GENERATED
6 November 2026 at 11:59 UTC

ASSIGNMENT ID
af711158-2501-46f6-b0d7-f96163c00e14

PROTOCOL ID
0297e174-2ac7-4702-a23e-6029a5b85a08

1. Study overview

EXPERIMENT TITLE
One-day offsite retreats, productivity and morale

RESEARCH QUESTION
Does inviting employees to a paid one-day offsite retreat, compared with a normal working day, raise their productivity, measured by a department-standardised weekly output index averaged over the eight weeks after the retreat date? Secondarily, does it raise job satisfaction four weeks later, and do any effects last to weeks 9-16 and the week 12 survey? The firm spends heavily on retreats but has never tested whether they pay off, and leadership wants evidence before offering one to all staff.

STUDY CONTEXT AND SETTING
A UK business-services company (fictional) with about 6,400 permanent employees in seven departments (Sales, Customer Service, Engineering, Operations, Finance, Marketing, HR & Administration) and four sites (London, Manchester, Birmingham, Glasgow). Many staff work hybrid and some are fully remote. Retreats run between 1 February and 5 March 2027 at a countryside conference venue within about 90 minutes of each site. Each department already records weekly output in its operational systems and the firm runs short staff surveys, so outcomes come from existing records plus two follow-up surveys. Practical constraints: the budget covers 500 retreat places, managers must release staff for a day, and February illness and the end-of-March quarter-end affect workloads.

2. Population and sample

TARGET POPULATION

Permanent employees of UK business-services firms that have several departments and a mix of office-based, hybrid and remote working.

ELIGIBILITY CRITERIA

Eligible: permanent employees on the HR roster on 4 January 2027 with at least three months' service, contracted for at least 0.5 FTE, and whose department records a weekly output metric. Excluded: directors and other senior leaders, contractors and agency staff, employees on long-term leave or with a notified leaving date at sampling, the HR staff running the study, and employees who opted out of the study's use of their records before sampling.

ACCESSIBLE POPULATION AND SAMPLING FRAME

About 5,900 eligible employees across all seven departments and four sites, taken from the HR roster extract of 4 January 2027 after opt-outs are removed. The study sample is a simple random sample of 1,000 employees drawn by computer from this roster.

RECRUITMENT AND CONSENT

In December 2026 every eligible employee receives an all-staff email explaining the study: that some employees will be invited to a retreat in February or March and others offered the same retreat after the study ends, which work records will be used, and how to opt out of the use of their records before 4 January. After sampling, the 1,000 sampled employees are told they are in the study. Consent to each survey is asked on its first screen, with up to two reminders per survey. Retreat-arm employees receive a calendar invitation for their assigned date and their manager is notified; travel is paid and the day counts as working time. There is no other incentive.

INTENDED SAMPLE SIZE

1000

SAMPLE-SIZE JUSTIFICATION

The venue budget funds 10 retreat sessions of 50 places, so 500 employees can be treated; an equal control arm gives 1,000. With 500 per arm, 80% power and a two-sided 5% test, the minimum detectable effect on output is about 0.18 standard deviations without covariates. Adjusting for baseline output, which should explain roughly half of the outcome variance, lowers it to about 0.13 SD (around 3 to 4 index points). For job satisfaction, with about 80% survey response, the minimum detectable effect is about 0.2 SD. Effects smaller than these would not justify a firm-wide roll-out.

3. Units and dataset structure

UNIT OF ASSIGNMENT AND INTERVENTION

Employee

UNIT OF OBSERVATION

Employee: weekly output from department systems, HR and payroll records, collaboration metadata and staff surveys

UNIT OF ANALYSIS

Employee

GROUPING OR REPEATED-OBSERVATION STRUCTURE

Employees are nested in teams (about 420 teams; most contain 1 to 5 sampled employees) within seven departments and four sites. Teammates share a line manager and workloads, and retreat and control employees

can be in the same team. Retreat-arm employees attend one of 10 sessions of about 50 people from mixed departments, so treated employees share a session. Each employee has baseline and several follow-up measurements, stored as separate columns.

EXPECTED ROW STRUCTURE

One row per sampled employee (1,000 rows), with baseline measures, assignment, attendance and follow-up measures in separate columns (wide format).

OBSERVATION SCHEDULE

26 Oct to 18 Dec 2026: baseline weekly output and contacts (8 working weeks).

4 to 15 Jan 2027: HR and payroll extract; baseline survey.

18 Jan 2027: sample of 1,000 drawn; arm and reference date assigned.

2 Feb to 4 Mar 2027: retreats on Tuesdays and Thursdays; the reference date is week 0 for both arms.

Weeks 1 to 8: weekly output and cross-department contacts. Week 4: survey 1.

Weeks 9 to 16: weekly output. Week 12: survey 2.

End of week 16: HR leaver records; study ends.

4. Treatments and conditions

Condition 1: Control

NAME

Control

CONDITION CONTENT

Business as usual. The employee works normally on their assigned reference date and receives no retreat invitation or retreat content during the 16-week study. Control employees complete the same surveys and are offered the same retreat once follow-up ends (waitlist control).

CONTROL OR COMPARISON CONDITION

Yes

START, DURATION, AND END

Works normally on the same assigned reference date (2 Feb to 4 Mar 2027); offered the retreat after week 16

Condition 2: Retreat

NAME

Retreat

CONDITION CONTENT

A paid one-day offsite retreat at a countryside conference venue within about 90 minutes of the employee's site, with travel provided and the day counted as working time. About 50 employees from mixed departments attend each of 10 sessions. Programme: a facilitated morning session on the firm's priorities and each person's part in them; a cross-department problem-solving workshop in groups of six; an outdoor team activity; lunch and an informal social hour; and a closing session on workload and wellbeing practices. No work is done that day and managers arrange cover. The programme and facilitation team are the same at every session.

CONTROL OR COMPARISON CONDITION

No

START, DURATION, AND END

One working day (09:00 to 17:00) on the employee's assigned date, between 2 Feb and 4 Mar 2027

5. Randomization

METHOD

stratified

ALLOCATION

1:1 within each stratum (about 500 Retreat and 500 Control)

STRATIFICATION VARIABLES

Department (7) x work arrangement (office-based, hybrid, remote): 21 strata. Permuted blocks of 4 within each stratum; when a stratum's size is not a multiple of 4, the last incomplete block is assigned at random, so arm sizes within a stratum differ by at most one and the overall split is as close to 500/500 as the strata allow. Within each arm, the reference date is then randomly assigned so both arms are spread evenly across the 10 session dates.

CONCEALMENT OR BLINDING

Blinding is not possible for employees or their managers. The allocation list is generated with a fixed seed by a statistician outside HR and the business units, only after the baseline data are frozen, so nobody can influence who is assigned. Output comes from operational systems rather than manager ratings, and the analyst is kept blind to arm labels until the analysis code is locked.

6. Complete variable inventory

Variable 1: employee_id

NAME

employee_id

ROLE

identifier

SCALE

E0001 to E1000; unique; linked to staff numbers only through the separate key

SOURCE

Generated by the study team at sampling

TIMING

Before randomization

DATA TYPE

text

FREQUENCY

Once, 18 January 2027

DERIVATION

Not provided

DESCRIPTION

Study identifier for each sampled employee. The study team keeps a separate, access-controlled key linking each ID to the HR staff number; it is used only to send invitations and merge HR, operational,

collaboration and survey records, and is never part of the analysis file

MISSINGNESS
Must not be missing

VALID VALUES
E0001 to E1000; unique; linked to staff numbers only through the separate key

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 18 January 2027

Variable 2: department

NAME
department

ROLE
baseline_covariate

SCALE
Sales; Customer Service; Engineering; Operations; Finance; Marketing; HR & Administration

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
categorical

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Department the employee works in; a stratification variable

MISSINGNESS
Must not be missing

VALID VALUES
Sales; Customer Service; Engineering; Operations; Finance; Marketing; HR & Administration

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 3: team_id

NAME
team_id

ROLE
cluster_group

SCALE
T001 to T420

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
categorical

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Anonymous code for the employee's team (everyone reporting to the same line manager)

MISSINGNESS
Must not be missing

VALID VALUES
T001 to T420

DATASET LEVEL
Team

COLLECTION MODE
direct

MEASURED ENTITY
Team

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 4: site

NAME
site

ROLE
baseline_covariate

SCALE
London; Manchester; Birmingham; Glasgow

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
categorical

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Office site the employee is attached to; remote employees are still attached to a site

MISSINGNESS
Must not be missing

VALID VALUES
London; Manchester; Birmingham; Glasgow

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 5: work_arrangement

NAME
work_arrangement

ROLE
moderator

SCALE
Office-based; Hybrid; Remote

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
categorical

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Contracted work arrangement; a stratification variable and a planned moderator of the retreat effect

MISSINGNESS
Must not be missing

VALID VALUES
Office-based; Hybrid; Remote

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 6: job_level

NAME
job_level

ROLE
baseline_covariate

SCALE
1 = Associate; 2 = Senior associate; 3 = Team lead; 4 = Manager

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
ordinal

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Job grade, listed from lowest to highest

MISSINGNESS
Must not be missing

VALID VALUES
1 = Associate; 2 = Senior associate; 3 = Team lead; 4 = Manager

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 7: tenure_years

NAME
tenure_years

ROLE
baseline_covariate

SCALE
0.25 to 40 years; one decimal place

SOURCE
HR record of employment start date

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, 4 January 2027 HR extract

DERIVATION
Days from the employment start date to 4 January 2027, divided by 365.25

DESCRIPTION
Years of service at the firm on 4 January 2027

MISSINGNESS
Must not be missing

VALID VALUES
0.25 to 40 years; one decimal place

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 HR extract

Variable 8: experience_years

NAME
experience_years

ROLE
baseline_covariate

SCALE
0 to 50 years; one decimal place; should be at least tenure_years

SOURCE
HR record (self-reported at hiring and updated at annual reviews)

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, 4 January 2027 HR extract

DERIVATION
Not provided

DESCRIPTION
Total years of paid work experience, including previous employers

MISSINGNESS
Missing for about 5% of employees whose HR record was never completed

VALID VALUES
0 to 50 years; one decimal place; should be at least tenure_years

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 HR extract

Variable 9: age

NAME
age

ROLE
baseline_covariate

SCALE
18 to 67 years; whole years

SOURCE
HR record (date of birth converted to age; the date of birth itself is not kept)

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, 4 January 2027 HR extract

DERIVATION
Completed years between the date of birth and 4 January 2027

DESCRIPTION
Age in completed years on 4 January 2027

MISSINGNESS
Must not be missing

VALID VALUES
18 to 67 years; whole years

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 HR extract

Variable 10: gender

NAME
gender

ROLE
baseline_covariate

SCALE
Woman; Man; Non-binary or self-described; Prefer not to say

SOURCE
HR equality-monitoring record

TIMING
Before randomization

DATA TYPE
categorical

FREQUENCY
Once, 4 January 2027 HR extract

DERIVATION
Not provided

DESCRIPTION
Gender as recorded in the employee's self-reported equality-monitoring record

MISSINGNESS
Not missing; Prefer not to say is a valid answer

VALID VALUES
Woman; Man; Non-binary or self-described; Prefer not to say

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 HR extract

Variable 11: education

NAME
education

ROLE
baseline_covariate

SCALE
1 = Secondary school; 2 = Vocational or apprenticeship; 3 = Undergraduate degree; 4 = Postgraduate degree

SOURCE
HR record (self-reported)

TIMING
Before randomization

DATA TYPE
ordinal

FREQUENCY
Once, 4 January 2027 HR extract

DERIVATION
Not provided

DESCRIPTION
Highest completed qualification, listed from lowest to highest

MISSINGNESS
Missing for about 8% of employees who never reported it

VALID VALUES
1 = Secondary school; 2 = Vocational or apprenticeship; 3 = Undergraduate degree; 4 = Postgraduate degree

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 HR extract

Variable 12: caring_responsibilities

NAME
caring_responsibilities

ROLE
baseline_covariate

SCALE
1 = yes; 0 = no

SOURCE
Baseline staff survey

TIMING
Before randomization

DATA TYPE
binary

FREQUENCY
Once, baseline survey (4 to 15 January 2027)

DERIVATION
Not provided

DESCRIPTION
Has regular caring responsibilities for a child or an adult dependant

MISSINGNESS
Missing if the baseline survey was not answered (about 10%)

VALID VALUES
1 = yes; 0 = no

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, baseline survey (4 to 15 January 2027)

Variable 13: commute_minutes

NAME
commute_minutes

ROLE
baseline_covariate

SCALE
0 to 150 minutes; self-reported and rounded to 5 minutes; 0 for remote employees

SOURCE
Baseline staff survey

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, baseline survey (4 to 15 January 2027)

DERIVATION
Not provided

DESCRIPTION
Usual one-way commute time to the site on office days

MISSINGNESS
Missing if the baseline survey was not answered (about 10%)

VALID VALUES
0 to 150 minutes; self-reported and rounded to 5 minutes; 0 for remote employees

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, baseline survey (4 to 15 January 2027)

Variable 14: full_time

NAME
full_time

ROLE
baseline_covariate

SCALE
1 = full time; 0 = part time

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
binary

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Contracted full time (1.0 FTE) rather than part time (0.5 to 0.9 FTE)

MISSINGNESS
Must not be missing

VALID VALUES
1 = full time; 0 = part time

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 15: annual_salary_gbp

NAME
annual_salary_gbp

ROLE
baseline_covariate

SCALE
18000 to 150000 pounds; whole pounds. A few records keyed in thousands by mistake appear as values below 1000 (for example 42 for 42000)

SOURCE
Payroll record

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, January 2027 payroll

DERIVATION
Not provided

DESCRIPTION
Full-time-equivalent base annual salary in pounds

MISSINGNESS
Must not be missing

VALID VALUES
18000 to 150000 pounds; whole pounds. A few records keyed in thousands by mistake appear as values below 1000 (for example 42 for 42000)

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, January 2027 payroll

Variable 16: team_size

NAME
team_size

ROLE
baseline_covariate

SCALE
2 to 25 people

SOURCE
HR roster extract

TIMING
Before randomization

DATA TYPE
count

FREQUENCY
Once, 4 January 2027 roster extract

DERIVATION
Not provided

DESCRIPTION
Number of people who report to the employee's line manager, including the employee

MISSINGNESS
Must not be missing

VALID VALUES
2 to 25 people

DATASET LEVEL
Team

COLLECTION MODE
direct

MEASURED ENTITY
Team

COLLECTION WINDOW
Once, 4 January 2027 roster extract

Variable 17: prior_performance_rating

NAME
prior_performance_rating

ROLE
baseline_covariate

SCALE
1 = Below expectations; 2 = Partly meets; 3 = Meets; 4 = Exceeds; 5 = Outstanding

SOURCE
Performance management system

TIMING
Before randomization

DATA TYPE
ordinal

FREQUENCY
Once, from the 2026 review cycle

DERIVATION
Not provided

DESCRIPTION
Rating from the 2026 annual performance review

MISSINGNESS
Missing for employees hired after 1 April 2026, who had no annual review (about 8%)

VALID VALUES
1 = Below expectations; 2 = Partly meets; 3 = Meets; 4 = Exceeds; 5 = Outstanding

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, from the 2026 review cycle

Variable 18: sick_days_prior_12m

NAME
sick_days_prior_12m

ROLE
baseline_covariate

SCALE
0 to 60 days; whole days

SOURCE
HR absence system

TIMING
Before randomization

DATA TYPE
count

FREQUENCY
Once, covering January to December 2026

DERIVATION
Sum of sickness-absence days recorded in the 12 months; employees with less than 12 months' service are counted over their time employed

DESCRIPTION
Days of recorded sickness absence in the 12 months before 4 January 2027

MISSINGNESS
Must not be missing

VALID VALUES
0 to 60 days; whole days

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, covering January to December 2026

Variable 19: baseline_cross_team_contacts

NAME
baseline_cross_team_contacts

ROLE
baseline_covariate

SCALE
0 to 200 colleagues

SOURCE
Aggregated collaboration-platform metadata (counts only; no message content)

TIMING
Before randomization

DATA TYPE
count

FREQUENCY
Counted once over the 8-week baseline window (26 October to 18 December 2026)

DERIVATION
Count of distinct colleague IDs in other departments across emails, chats and meetings in the window

DESCRIPTION
Number of distinct colleagues outside the employee's own department with whom they exchanged at least one work message or shared a meeting during the 8 baseline weeks

MISSINGNESS
Must not be missing

VALID VALUES
0 to 200 colleagues

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Counted once over the 8-week baseline window (26 October to 18 December 2026)

Variable 20: baseline_output_index

NAME
baseline_output_index

ROLE
baseline_covariate

SCALE
0 to 400 index points; one decimal place

SOURCE
Department operational systems: Sales CRM (value of closed orders), Customer Service ticketing (tickets resolved to quality standard), Engineering tracker (accepted story points), Operations (orders processed), Finance (reconciliations and reports completed), Marketing (approved deliverables), HR & Administration (cases closed)

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Weekly for 8 weeks (26 October to 18 December 2026); averaged

DERIVATION
Each week: the employee's value on the department's core metric, divided by their FTE, divided by the department's fixed reference (the median FTE-adjusted weekly output across all of the department's employee-weeks in the 8 baseline weeks), times 100. Then averaged over the weeks the employee worked

DESCRIPTION
Mean weekly output index over the 8 baseline weeks. 100 = the department's fixed reference: the median full-time-equivalent weekly output across all of that department's employee-weeks in the baseline period

MISSINGNESS
Missing if the employee worked fewer than 2 of the 8 baseline weeks

VALID VALUES
0 to 400 index points; one decimal place

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Weekly for 8 weeks (26 October to 18 December 2026); averaged

Variable 21: job_satisfaction_baseline

NAME
job_satisfaction_baseline

ROLE
baseline_covariate

SCALE
1 to 7 in steps of 0.2; higher = more satisfied

SOURCE
Baseline staff survey

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, baseline survey (4 to 15 January 2027)

DERIVATION
Mean of the five items; computed only when all five are answered

DESCRIPTION
Job satisfaction: mean of five 1-to-7 agreement items (satisfied with the job overall, the work itself, colleagues, line manager, and feeling valued)

MISSINGNESS
Missing if the baseline survey was not answered (about 10%)

VALID VALUES
1 to 7 in steps of 0.2; higher = more satisfied

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, baseline survey (4 to 15 January 2027)

Variable 22: belonging_baseline

NAME
belonging_baseline

ROLE
moderator

SCALE
1 to 7 in steps of 0.25; higher = stronger belonging

SOURCE
Baseline staff survey

TIMING
Before randomization

DATA TYPE
continuous

FREQUENCY
Once, baseline survey (4 to 15 January 2027)

DERIVATION
Mean of the four items; computed only when all four are answered

DESCRIPTION
Sense of belonging and connection to colleagues: mean of four 1-to-7 agreement items (for example, 'I feel part of a community at work'); a planned moderator

MISSINGNESS
Missing if the baseline survey was not answered (about 10%)

VALID VALUES
1 to 7 in steps of 0.25; higher = stronger belonging

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, baseline survey (4 to 15 January 2027)

Variable 23: retreat_arm

NAME
retreat_arm

ROLE
treatment_assignment

SCALE
Retreat; Control

SOURCE
Randomization list

TIMING
At randomization

DATA TYPE
categorical

FREQUENCY
Once, 18 January 2027

DERIVATION
Not provided

DESCRIPTION
Randomly assigned condition

MISSINGNESS
Must not be missing

VALID VALUES
Retreat; Control

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 18 January 2027

Variable 24: session_date

NAME
session_date

ROLE
time

SCALE
One of 10 dates (YYYY-MM-DD): 2027-02-02; 2027-02-04; 2027-02-09; 2027-02-11; 2027-02-16; 2027-02-18; 2027-02-23; 2027-02-25; 2027-03-02; 2027-03-04

SOURCE
Randomization list

TIMING
At randomization

DATA TYPE
date_time

FREQUENCY
Once, 18 January 2027

DERIVATION
Not provided

DESCRIPTION
Reference date: the retreat session date assigned to the employee at randomization. Retreat-arm employees are invited to the retreat on this date (attended_retreat records whether they actually went); Control employees work normally. It is week 0 for every follow-up window, whether or not the employee attended

MISSINGNESS
Must not be missing

VALID VALUES
One of 10 dates (YYYY-MM-DD): 2027-02-02; 2027-02-04; 2027-02-09; 2027-02-11; 2027-02-16; 2027-02-18; 2027-02-23; 2027-02-25; 2027-03-02; 2027-03-04

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, 18 January 2027

Variable 25: attended_retreat

NAME
attended_retreat

ROLE
compliance_measure

SCALE
1 = attended; 0 = did not attend; always 0 in the Control arm

SOURCE
Venue sign-in sheet, entered by HR

TIMING
Week 0 (reference date)

DATA TYPE
binary

FREQUENCY
Once

DERIVATION
Not provided

DESCRIPTION
Whether the employee attended the retreat on their assigned date (signed in at the venue)

MISSINGNESS
Must not be missing

VALID VALUES
1 = attended; 0 = did not attend; always 0 in the Control arm

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once

Variable 26: cross_team_contacts_w1_8

NAME
cross_team_contacts_w1_8

ROLE
mediator

SCALE
0 to 200 colleagues

SOURCE
Aggregated collaboration-platform metadata (counts only; no message content)

TIMING
After treatment, weeks 1 to 8

DATA TYPE
count

FREQUENCY
Counted once over the 8-week window

DERIVATION
Count of distinct colleague IDs in other departments across emails, chats and meetings in the window

DESCRIPTION
Number of distinct colleagues outside the employee's own department with whom they exchanged at least one work message or shared a meeting in weeks 1 to 8 after the reference date

MISSINGNESS
Missing if the employee left the firm before the end of week 8

VALID VALUES
0 to 200 colleagues

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Counted once over the 8-week window

Variable 27: output_index_w1_8

NAME
output_index_w1_8

ROLE
outcome

SCALE
0 to 400 index points; one decimal place

SOURCE
Department operational systems (the same metrics as baseline_output_index)

TIMING
After treatment, weeks 1 to 8

DATA TYPE
continuous

FREQUENCY
Weekly for 8 weeks; averaged

DERIVATION
Each week: the employee's value on the department's core metric, divided by their FTE, divided by the same fixed department reference used for baseline_output_index, times 100. Then averaged over the weeks the employee worked in weeks 1 to 8

DESCRIPTION
Primary productivity outcome: mean weekly output index over weeks 1 to 8 after the reference date, on the same fixed department reference as baseline_output_index (100 = the department's baseline median FTE-adjusted weekly output)

MISSINGNESS
Missing if the employee worked fewer than 2 of the 8 weeks (left the firm or long leave)

VALID VALUES
0 to 400 index points; one decimal place

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Weekly for 8 weeks; averaged

Variable 28: output_index_w9_16

NAME
output_index_w9_16

ROLE
outcome

SCALE
0 to 400 index points; one decimal place

SOURCE
Department operational systems (the same metrics as baseline_output_index)

TIMING
After treatment, weeks 9 to 16

DATA TYPE
continuous

FREQUENCY
Weekly for 8 weeks; averaged

DERIVATION
Each week: the employee's value on the department's core metric, divided by their FTE, divided by the same fixed department reference used for baseline_output_index, times 100. Then averaged over the weeks the employee worked in weeks 9 to 16

DESCRIPTION
Persistence outcome: mean weekly output index over weeks 9 to 16 after the reference date, on the same fixed department reference as baseline_output_index and output_index_w1_8

MISSINGNESS
Missing if the employee worked fewer than 2 of the 8 weeks (left the firm or long leave)

VALID VALUES
0 to 400 index points; one decimal place

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Weekly for 8 weeks; averaged

Variable 29: job_satisfaction_w4

NAME
job_satisfaction_w4

ROLE
outcome

SCALE
1 to 7 in steps of 0.2; higher = more satisfied

SOURCE
Follow-up staff survey 1

TIMING
After treatment, week 4

DATA TYPE
continuous

FREQUENCY
Once; survey open for 7 days

DERIVATION
Mean of the five items; computed only when all five are answered

DESCRIPTION
Job satisfaction four weeks after the reference date, using the same five items as job_satisfaction_baseline

MISSINGNESS
Missing if the survey was not answered (about 20%) or the employee had left the firm

VALID VALUES
1 to 7 in steps of 0.2; higher = more satisfied

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once; survey open for 7 days

Variable 30: job_satisfaction_w12

NAME
job_satisfaction_w12

ROLE
outcome

SCALE
1 to 7 in steps of 0.2; higher = more satisfied

SOURCE
Follow-up staff survey 2

TIMING
After treatment, week 12

DATA TYPE
continuous

FREQUENCY
Once; survey open for 7 days

DERIVATION
Mean of the five items; computed only when all five are answered

DESCRIPTION
Job satisfaction twelve weeks after the reference date, using the same five items as job_satisfaction_baseline

MISSINGNESS
Missing if the survey was not answered (about 30%) or the employee had left the firm

VALID VALUES
1 to 7 in steps of 0.2; higher = more satisfied

DATASET LEVEL
Employee

COLLECTION MODE
derived

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once; survey open for 7 days

Variable 31: left_firm_w16

NAME
left_firm_w16

ROLE
outcome

SCALE
1 = left; 0 = still employed

SOURCE
HR leaver records

TIMING
After treatment, by week 16

DATA TYPE
binary

FREQUENCY
Once, at the end of week 16

DERIVATION
Not provided

DESCRIPTION
Whether the employee left the firm for any reason between the reference date and the end of week 16

MISSINGNESS
Must not be missing

VALID VALUES
1 = left; 0 = still employed

DATASET LEVEL
Employee

COLLECTION MODE
direct

MEASURED ENTITY
Employee

COLLECTION WINDOW
Once, at the end of week 16

7. Outcomes

Outcome 1: output_index_w1_8

LINKED VARIABLE
output_index_w1_8

LEVEL
Employee

PRIORITY
primary

COMPARISON
Retreat arm versus Control arm, as randomized (intention to treat)

AGGREGATION
Mean of the weekly output index over the weeks worked in weeks 1 to 8; one value per employee

EVENT STRUCTURE
repeated

EXPOSURE PERIOD
Not provided

MEASUREMENT WINDOW
Weeks 1 to 8 after the reference date

Outcome 2: job_satisfaction_w4

LINKED VARIABLE
job_satisfaction_w4

LEVEL
Employee

PRIORITY
secondary

COMPARISON
Retreat arm versus Control arm, as randomized (intention to treat)

AGGREGATION
Mean of the five survey items; one value per employee

EVENT STRUCTURE
once

EXPOSURE PERIOD
Not provided

MEASUREMENT WINDOW
Week 4 survey after the reference date (open for 7 days)

Outcome 3: output_index_w9_16

LINKED VARIABLE
output_index_w9_16

LEVEL
Employee

PRIORITY
secondary

COMPARISON
Retreat arm versus Control arm, as randomized (intention to treat)

AGGREGATION
Mean of the weekly output index over the weeks worked in weeks 9 to 16; one value per employee

EVENT STRUCTURE
repeated

EXPOSURE PERIOD
Not provided

MEASUREMENT WINDOW
Weeks 9 to 16 after the reference date

Outcome 4: job_satisfaction_w12

LINKED VARIABLE
job_satisfaction_w12

LEVEL
Employee

PRIORITY
secondary

COMPARISON
Retreat arm versus Control arm, as randomized (intention to treat)

AGGREGATION
Mean of the five survey items; one value per employee

EVENT STRUCTURE
once

EXPOSURE PERIOD
Not provided

MEASUREMENT WINDOW
Week 12 survey after the reference date (open for 7 days)

Outcome 5: left_firm_w16

LINKED VARIABLE
left_firm_w16

LEVEL
Employee

PRIORITY
secondary

COMPARISON
Share who left: Retreat arm versus Control arm, as randomized (intention to treat)

AGGREGATION
One indicator per employee: 1 if an exit is recorded by the end of week 16

EVENT STRUCTURE
once

EXPOSURE PERIOD
Not provided

MEASUREMENT WINDOW
From the reference date to the end of week 16

8. Hypotheses

Hypothesis 1

LINKED OUTCOME
output_index_w1_8

STATEMENT
Employees assigned to the retreat will have a higher mean weekly output index in weeks 1 to 8 after the reference date than employees assigned to a normal working day, because the retreat strengthens working relationships across departments and renews motivation.

PREDICTED DIRECTION
Higher

Hypothesis 2

LINKED OUTCOME
job_satisfaction_w4

STATEMENT
Employees assigned to the retreat will report higher job satisfaction four weeks after the reference date than employees assigned to a normal working day.

PREDICTED DIRECTION
Higher

Hypothesis 3

LINKED OUTCOME
output_index_w9_16

STATEMENT
Exploratory: it is unclear whether any productivity difference lasts into weeks 9 to 16; the effect of a single day may fade as the retreat recedes.

PREDICTED DIRECTION
Uncertain (exploratory)

Hypothesis 4

LINKED OUTCOME
job_satisfaction_w12

STATEMENT
Exploratory: it is unclear whether any difference in job satisfaction remains at the week 12 survey.

PREDICTED DIRECTION
Uncertain (exploratory)

Hypothesis 5

LINKED OUTCOME
left_firm_w16

STATEMENT
Employees assigned to the retreat will be less likely to have left the firm by the end of week 16 than employees assigned to a normal working day, because a stronger sense of belonging makes leaving less attractive.

PREDICTED DIRECTION
Lower

9. Analysis plan

Main analyses

HYPOTHESIS 1: PLANNED ANALYSIS
Intention to treat: employees are analysed in the arm they were assigned to, whether or not they attended. The main analysis population is every randomized employee with an observed output_index_w1_8 (about 95% are expected); the sensitivity checks under Handling missing data extend the comparison to all 1,000 randomized employees. Estimate the effect as the coefficient on a Retreat indicator in an OLS regression of output_index_w1_8 on the indicator, stratum fixed effects (department x work arrangement), baseline_output_index and session_date fixed effects (calendar timing). Report the estimate in index points and in standard-deviation units with a 95% confidence interval, and judge H1 by a two-sided test at the 5% level. Also report the unadjusted difference in means.

HYPOTHESIS 2: PLANNED ANALYSIS
Same intention-to-treat regression with job_satisfaction_w4 as the outcome and job_satisfaction_baseline in place of baseline output (missing baseline values filled with the mean plus a missing indicator). Report the difference in scale points and standard-deviation units with a 95% confidence interval; two-sided test at 5%, Holm-adjusted across the secondary family.

HYPOTHESIS 3: PLANNED ANALYSIS
Same model as H1 with output_index_w9_16 as the outcome. Compare the weeks 9 to 16 estimate with the weeks 1 to 8 estimate in a stacked regression (both windows, errors clustered by employee and team) to judge fade-out. No direction is tested; report the estimate and 95% confidence interval.

HYPOTHESIS 4: PLANNED ANALYSIS
Same model as H2 with job_satisfaction_w12 as the outcome; report the estimate and 95% confidence interval and compare it with the week 4 estimate. Exploratory; no direction is tested.

HYPOTHESIS 5: PLANNED ANALYSIS
Linear probability model of left_firm_w16 on the Retreat indicator with stratum fixed effects, job_satisfaction_baseline and tenure_years; report the difference in percentage points with a 95% confidence interval and check it with a logistic regression. Two-sided test, Holm-adjusted across the secondary family.

ADDITIONAL ANALYSES
(1) Balance table of baseline covariates by arm. (2) Effect of attending: complier average causal effect by two-stage least squares, using random assignment as the instrument for attended_retreat. (3) Pre-specified moderators, each as an interaction with the Retreat indicator: work_arrangement, belonging_baseline (below versus above the median), tenure under 2 years, and job_level. (4) Mediation, descriptive only: the effect on cross_team_contacts_w1_8 and whether adding it shrinks the output effect. (5) Spillovers: whether Control employees with a larger share of sampled teammates in the Retreat arm have different outcomes. (6) Variation in effects across the 10 retreat sessions. (7) Robustness: trimming the top 1% of output values at the same pooled cut-off in both arms, and using log(1 + output index) so zero weeks are kept.

HANDLING MISSING DATA
The main analyses use every randomized employee with an observed value of the outcome, analysed in their assigned arm. Report missingness by arm and test whether it is related to arm and baseline characteristics. Sensitivity checks covering all 1,000 randomized employees: (a) multiple imputation (20 imputations) from baseline covariates and arm, which assumes outcomes are missing at random given those variables; (b) Lee

bounds, which need no such assumption but are wider, if attrition differs by arm. left_firm_w16 should have no missing values. Missing baseline covariates are filled with the sample mean plus a missing indicator, so nobody is dropped for a missing covariate.

UNCERTAINTY AND CLUSTERING

Employees are randomized individually within strata, but teammates share a manager and workload and treated employees share a retreat session, so the main standard errors are clustered by team (about 420 clusters). As checks: two-way clustering by team and by session (Control employees grouped by their reference date), heteroskedasticity-robust standard errors, and randomization-inference p-values from 2,000 re-runs of the actual allocation procedure (stratified permuted blocks, then random reference dates).

MULTIPLE COMPARISONS

H1 is the single primary hypothesis, tested at 5%. H2 and H5, the directional secondary hypotheses, form one family with a Holm adjustment. H3 and H4 are exploratory persistence questions: they are reported as estimates with 95% confidence intervals, with unadjusted p-values for information only, and no confirmatory claim is made from them. Moderator, mediation, spillover and session analyses are also exploratory and labelled as such.

10. Expected data-generating context

CAUSAL DIAGRAM CODE

retreat_arm -> attended_retreat
attended_retreat -> cross_team_contacts_w1_8
attended_retreat -> job_satisfaction_w4
attended_retreat -> output_index_w1_8
attended_retreat -> left_firm_w16
cross_team_contacts_w1_8 -> output_index_w1_8
job_satisfaction_w4 -> job_satisfaction_w12
output_index_w1_8 -> output_index_w9_16
caring_responsibilities -> attended_retreat
full_time -> attended_retreat
work_arrangement -> belonging_baseline
belonging_baseline -> job_satisfaction_baseline
job_satisfaction_baseline -> job_satisfaction_w4
job_satisfaction_baseline -> left_firm_w16
tenure_years -> left_firm_w16
baseline_output_index -> output_index_w1_8
baseline_output_index -> output_index_w9_16
prior_performance_rating -> baseline_output_index
tenure_years -> baseline_output_index
baseline_cross_team_contacts -> cross_team_contacts_w1_8
job_level -> annual_salary_gbp
tenure_years -> annual_salary_gbp
site -> annual_salary_gbp
age -> experience_years

EXPECTED COVARIATE RELATIONSHIPS

Age, experience and tenure rise together, and experience is normally at least as long as tenure. Salary rises with job level and tenure, is higher in London, and is higher in Engineering and Sales. Prior performance ratings are positively related to baseline output (correlation around 0.4), and baseline output strongly predicts later output (correlation around 0.6 to 0.7). Remote employees have lower baseline belonging and fewer cross-department contacts; belonging is positively related to job satisfaction. Low satisfaction and short tenure go with more sick days and a higher chance of leaving. Caring responsibilities are most common at ages 30 to 50 and go with part-time contracts. Remote employees have a commute of 0. Team leads and managers have more cross-department contacts. Randomization should balance all of these between arms.

MODERATORS OR TREATMENT-EFFECT DIFFERENCES

Effects on belonging, job satisfaction and output may be larger for remote and hybrid employees, for employees with low baseline belonging, and for employees with under two years' tenure, and smaller for managers. Output effects may be smaller in Sales, where output depends on deal timing. Effects may fade between weeks 1 to 8 and weeks 9 to 16, and between the week 4 and week 12 surveys.

COMPLIANCE OR DELIVERY PROBLEMS

One-sided non-compliance: roughly 10 to 15% of Retreat-arm employees may not attend because of illness, client deadlines, caring responsibilities or travel distance, more often among employees with caring responsibilities, part-time staff, and Sales staff near the quarter-end. Control employees cannot attend because HR controls the venue list, so attended_retreat is 0 for every Control employee. A few attendees may leave early; this is not recorded.

MISSING-DATA AND ATTRITION PROCESS

Employees who leave the firm stop contributing output and surveys: about 3 to 5% by week 8 and 7 to 10% by week 16, more often those with short tenure and low baseline satisfaction. Survey non-response is about 10% at baseline, 20% at week 4 and 30% at week 12; less satisfied employees may be less likely to respond, and retreat attendees may feel more obliged to respond, so response may differ by arm. An output window is missing when the employee worked fewer than 2 of its 8 weeks (long sickness or parental leave, about 2%). Experience, education and prior rating have the small gaps described in the variable inventory.

MEASUREMENT LIMITATIONS AND UNUSUAL OBSERVATIONS

The output index is right-skewed and noisier in Sales, where single large deals create occasional very high weeks (index above 300); it cannot fall below 0. Engineering story points can drift upwards over time. Satisfaction and belonging scores are bounded at 1 and 7, with some ceiling effects among very satisfied employees. Commute times are self-reported and rounded to 5 minutes. As noted in the variable inventory, a few annual_salary_gbp records were keyed in thousands and appear as values below 1000 (for example 42 for 42000); these are data-entry errors to be corrected during cleaning, not real salaries. A few employees may have experience_years below tenure_years because of out-of-date HR records.

TIME AND DEPENDENCY PATTERNS

Outcomes are correlated within teams because teammates share a manager, workload and norms; treated employees in the same session share facilitators, weather and group mix, so outcomes may also cluster by session. Spillovers are possible: Control employees whose teammates attended may gain from new contacts or lose output while covering on the retreat day. Calendar effects: February illness lowers output for early sessions, and the end-of-March quarter-end raises Sales output in weeks 1 to 8 for later session dates. Weekly output is autocorrelated within an employee, so the two output windows are correlated.

11. Submitter responses and justifications

AI-R1
Agreed. The primary outcome tells us whether output is higher after the retreat, not whether the retreat pays for itself. In the report we will set any output gain against one lost working day per attendee and the venue cost in a simple break-even calculation, and we will not claim that a firm-wide roll-out pays off from the output effect alone.

AI-R2
Accepted. Averaging over weeks worked keeps output comparable for employees with some leave, but it can hide differences in time at work. We will report weeks worked by arm and, as a check, a version that treats weeks not worked as zero output, so readers can see whether the conclusion depends on this choice.

AI-R3
Accepted. Observed-outcome estimates apply to all randomized employees only if missingness is unrelated to outcomes within arm. We will show them alongside the missingness-by-arm table, multiple imputation (missing at random given baseline covariates) and Lee bounds (which assume assignment can only push response in one direction). If these disagree, the bounds are reported as the main statement of uncertainty.

AI-R4
We will interpret the intention-to-treat estimate as the effect of offering the retreat when about half of sampled staff are offered it, including any spillovers onto teammates. The attendance (CACE) estimate needs the exclusion restriction and no interference, so it is reported as indicative only. The spillover analysis is exploratory because most teams contain only a few sampled employees.

AI-R5
Agreed that 10 date clusters are too few for reliable cluster-robust errors. Team-clustered errors remain the main analysis; randomization inference that repeats the real allocation is the main check; the two-way clustered results are a sensitivity check only and will be flagged as possibly unstable.

AI-R6
Accepted. The stated minimum detectable effects are optimistic. Non-attendance of 10 to 15%, about 5% missing primary outcomes and modest team clustering could raise them by roughly a fifth, so we will report the achieved confidence-interval width and treat a non-significant result as inconclusive rather than as evidence of no effect.

AI-R7
Agreed. The mediation analysis will only describe whether contacts and output move together. We will not claim that contacts caused the output change, because both are measured over the same weeks and contacts are not randomized.

AI-R8
Accepted. Findings will be stated for eligible employees of this firm. We will discuss which features (hybrid working, department mix, the retreat format) might make other UK business-services firms similar or different, but we will not claim the effect transfers directly.

Document boundary

This appendix contains the approved design information supplied by the submitter. It does not disclose hidden effects, distribution parameters, simulation code, random seeds, missingness parameters, or anomaly rates.