capstract:2611.00026v1 [med.PH] 13 Nov 2026

Original research

Reply-all abstinence and work-related burnout in office-based employees: a pragmatic cluster-randomised controlled trial

Siobhan R. Achterberg[1,†], Oluwaseun T. Bakare[2], Mikkel H. Strand[1], Priya Natarajan-Holt[3], Gwendolyn A. Pyke[1,2]

[1] Department of Occupational and Environmental Medicine, Calder Vale University, Manchester, UK
[2] Northgate Centre for Trials and Statistics, Northgate Institute of Health Research, Leeds, UK
[3] School of Psychology, Wexcombe University, Cardiff, UK

Correspondence: Dr Siobhan R. Achterberg, Department of Occupational and Environmental Medicine, Calder Vale University, Manchester, UK (s.achterberg@caldervale.ac.uk)

Abstract

Background: Email volume is associated with work-related stress, and messages sent with the reply-all function make up a large share of the email that office workers receive. Whether reducing reply-all use lowers burnout has not been tested in a randomised trial.

Methods: We did a pragmatic, two-arm, cluster-randomised controlled trial in 48 teams from three organisations in England. Teams were randomly assigned (1:1, stratified by organisation) to a 12-week reply-all abstinence intervention (removal of the reply-all command from all email clients, a 30-minute group session and weekly feedback) or to usual practice with an attention-matched session on email security. The primary outcome was work-related burnout on the Copenhagen Burnout Inventory (CBI; range 0–100) at 12 weeks, analysed by intention to treat with a linear mixed model. Abstinence was verified from mail-server metadata. The trial is registered with the Occupational Interventions Trial Register, OITR-2025-00417.

Findings: Between 7 April and 30 May 2025, 671 employees in 48 teams were enrolled; 338 (24 teams) were allocated to the intervention and 333 (24 teams) to control, and 620 (92%) completed the 12-week assessment. Mean work-related burnout fell from 48.6 to 43.9 points in the intervention group and from 47.9 to 47.6 in the control group (adjusted mean difference −4.1, 95% CI −6.9 to −1.3; p=0.004). Inbound reply-all messages were 54% lower in intervention teams (rate ratio 0.46, 0.39–0.55). Continuous log-verified abstinence for 12 weeks was achieved by 42% of intervention participants and by 25% of those who had sent ten or more reply-all messages a week at baseline. More intervention participants reported feeling left out of the loop (31% *vs* 18%; odds ratio 2.03, 1.38–2.98). At 24 weeks, 12 weeks after the command was restored, the difference in burnout was −1.2 (−4.1 to 1.7). No serious adverse events were reported.

Interpretation: Team-level reply-all abstinence produced a small reduction in work-related burnout, below the 5-point difference the trial was designed to detect, which did not persist once the command was restored. Lapses were common, particularly among heavy users.

Funding: Workplace Health Research Fund.

Keywords: burnout; email; occupational health; cluster-randomised trial; workplace intervention

Introduction

Burnout is classified in ICD-11 as an occupational phenomenon resulting from chronic workplace stress that has not been successfully managed [1]. It is usually described in terms of exhaustion, cynicism towards the job and reduced professional efficacy [2], and it is associated with sickness absence and staff turnover. Interventions aimed at the organisation of work have generally produced larger effects than those aimed at individual coping, but both have been modest, and most trials have been in healthcare settings [3, 4].

Email is one of the more tractable features of office work. In a mixed-methods study of a technology company, time spent handling email predicted perceived overload, and email came to stand for the overload even when other demands were responsible [5]. Restricting how often people check email lowered daily stress in a within-person experiment [6], and removing email altogether for a working week lowered physiological markers of stress in a small field study [7]. Neither approach is practical for most organisations.

Reply-all is a narrower target. A metadata study of 1.4 million enterprise email threads found that replies addressed to all recipients of the previous message made up 11% of messages sent but 29% of messages received, because each one reaches many people, and that they were the main route by which threads grew beyond ten messages [8]. A reduction in reply-all use by one person benefits colleagues more than it benefits the sender, so we treated the team, not the individual, as the unit of intervention. In a

pilot in two teams, removing the reply-all command from email clients for six weeks was acceptable to staff and reduced inbound email by about a fifth [9].

Defining and verifying abstinence from a habitual behaviour is a familiar problem in smoking cessation research, and we adopted its conventions: a pre-specified definition of a lapse, objective verification, and continuous abstinence as the main measure of adherence [10, 11]. Here we report a pragmatic cluster-randomised trial of team-level reply-all abstinence, with work-related burnout as the primary outcome.

Methods

Study design

This was a pragmatic, two-arm, parallel-group, cluster-randomised controlled trial with 1:1 allocation, done in three organisations in England: a regional public-sector administrative body (organisation A), an engineering consultancy (B) and an insurance services company ©. Teams were the unit of randomisation because reply-all messages are, by definition, received by colleagues, and individual randomisation would have exposed control participants to the intervention through their teammates. The trial was registered before enrolment with the Occupational Interventions Trial Register (OITR-2025-00417, 3 March 2025), and the protocol and statistical analysis plan were finalised before the database was locked. We report it in accordance with the CONSORT extension for cluster randomised trials [12, 13].

Participants

Teams were eligible if they had between 6 and 30 members, used the organisation’s corporate email system as a main channel of internal communication, and were not expecting a restructuring within 6 months. Within eligible teams, employees were

eligible if they worked at least 0.6 full-time equivalent, used corporate email daily, and were not expecting more than 4 weeks of planned leave during the intervention period. Staff whose role required broadcast communication to large groups, such as internal communications officers, were excluded. Team managers were eligible and were included on the same terms as other members.

Teams were approached through the occupational health or human resources function of each organisation. After a manager had agreed to the team’s participation, all eligible members were invited to an information session and gave individual written informed consent online. Consent and baseline questionnaires were completed before allocation was disclosed.

Randomisation and masking

Teams were allocated 1:1 within strata defined by organisation, using permuted blocks of two and four generated by an independent statistician at the Northgate Centre for Trials and Statistics who had no contact with participating teams or with the trial statistician. Allocation was revealed to the research team for all teams in an organisation at the same time, after baseline data collection in that organisation had closed. Participants and managers could not be masked. Questionnaire outcomes were self-reported online, and email outcomes were derived automatically from server logs. The trial statistician remained masked to allocation until the primary analysis was complete.

Procedures

The intervention had three components. First, the organisation’s IT department applied an administrative policy that removed the reply-all command from every email client

used by members of the team, including mobile applications. Participants could still reply to the sender and add recipients by hand. Second, each team attended a 30-minute group session, delivered by a member of the research team, that explained the rationale and agreed a team convention for messages that needed a group response. Third, each participant received a weekly email, addressed to them alone, summarising the number of lapses (defined below) in their team in the previous week without identifying individuals. The intervention ran for 12 weeks, after which the command was restored.

Control teams continued to use email as usual and attended a 30-minute session on recognising phishing messages, matched in length and format to the intervention session. They received no feedback emails.

In the first 3 weeks the policy did not take effect on one mobile email application used in organisations A and C, so the reply-all command remained available to participants who used that application on their phones. The problem was identified from the server logs and corrected on day 18. Participants could ask to have the command restored on their own account at any time; those who did so remained in the trial and were analysed as allocated.

Outcomes

The primary outcome was work-related burnout at 12 weeks, measured with the seven-item work-related subscale of the Copenhagen Burnout Inventory (CBI), scored from 0 to 100 with higher scores indicating more burnout [14]. The CBI was also administered at 4, 8 and 24 weeks.

Secondary outcomes at 12 weeks were personal burnout (CBI personal subscale), work engagement on the nine-item Utrecht Work Engagement Scale (UWES-9; range 0–6) [15], sickness absence from human resources records, and four outcomes derived from mail-server metadata: inbound emails per working day, inbound reply-all messages per working day, emails sent outside working hours (before 0800, after 1900 or at weekends) per week, and, in the intervention group, continuous abstinence. A single questionnaire item asked participants whether, in the previous 4 weeks, they had felt left out of the loop on matters relevant to their work (five-point scale, dichotomised at agree or strongly agree). We also recorded incidents of missed information reported by participants, and serious adverse events.

Because the intervention removed the command but not the behaviour, a lapse was defined by the recipients of a message and not by the button used to send it. A reply r sent by participant i to a message m with recipient set $R(m)$, where $|R(m)| \geq 3$, was classified as a reply-all message if

$$\frac{|R(r) \cap R(m)|}{|R(m) \setminus \{i\}|} \geq 0.9. \tag{1}$$

The same definition was applied in both groups, so that reply-all use could be compared regardless of how the recipient list was assembled. Continuous abstinence was defined as no lapse between day 1 and day 84, with no grace period [11]. Log extraction was limited to metadata, with recipient lists hashed before transfer; no message content was accessed.

Statistical analysis

We assumed a standard deviation of 18 points for the CBI work-related subscale [14], an intracluster correlation coefficient (ICC) of 0.04 and a mean of 13 analysed participants per team. With cluster size m and ICC ρ, the design effect is

$$\mathrm{DE} = 1 + (m - 1)\rho, \tag{2}$$

which gives 1.48. An individually randomised trial would need 204 participants per arm to detect a difference of 5 points, which we regarded as the minimal important difference, with 80% power at a two-sided α of 0.05. Inflating by the design effect gives 302 per arm, and allowing for 10% attrition gives 336. At an expected 14 consenting participants per team this required 24 teams per arm [16].

The primary analysis was by intention to treat, including all randomised participants in the groups to which their teams were allocated. The 12-week CBI score Y_{ij} of participant i in team j was modelled as

$$Y_{ij} = \beta_0 + \beta_1 T_j + \beta_2 Y_{ij}^{(0)} + \sum_{k} \gamma_k S_{kj} + u_j + \varepsilon_{ij}, \qquad u_j \sim \mathcal{N}(0, \sigma_u^2),\ \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{3}$$

where T_j indicates allocation to the intervention, $Y_{ij}^{(0)}$ is the baseline score and S_{kj} are indicators for organisation, the stratification factor. β_1 is the adjusted mean difference. Degrees of freedom were obtained with the Kenward–Roger method [17], and the ICC was estimated as

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2}. \tag{4}$$

Missing 12-week scores were handled by multiple imputation with chained equations (50 imputations), with allocation, team, organisation, baseline variables and the 4-week

and 8-week scores in the imputation model [18]. Sensitivity analyses were restricted to complete cases or adjusted further for age, sex and managerial role.

Continuous secondary outcomes were analysed with models of the same form. Counts were analysed with negative binomial mixed models with a random intercept for team, the log of working days observed as an offset, and adjustment for the baseline count and organisation, and are reported as rate ratios. The binary outcome was analysed with a mixed logistic model. Time to first lapse in the intervention group was described with Kaplan–Meier curves [19] and compared between heavy users (ten or more reply-all messages sent per week at baseline) and light users with a Cox model with a shared frailty for team [20]; participants who left their organisation were censored at the date of leaving.

Pre-specified subgroup analyses examined baseline reply-all use, managerial role and organisation, by adding an interaction with allocation to equation (3). As an exploratory analysis we estimated the proportion of the effect on burnout mediated by the reduction in inbound reply-all messages, using the quasi-Bayesian approach of Imai and colleagues [21]. Analyses used R version 4.4.1. No interim analyses were planned or done.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Results

Between 7 April and 30 May 2025, 61 teams (1,047 staff) were assessed for eligibility and 48 teams were randomised (Figure 1). Of the 812 eligible staff in these teams, 671 (83%) consented and completed baseline assessments: 338 in 24 intervention teams and 333 in 24 control teams. Teams had a median of 14 participants (range 6–25). All 48 teams remained in the trial. At 12 weeks, 311 intervention and 309 control participants completed the primary outcome questionnaire (92% overall). The last 24-week follow-up was completed on 21 November 2025.

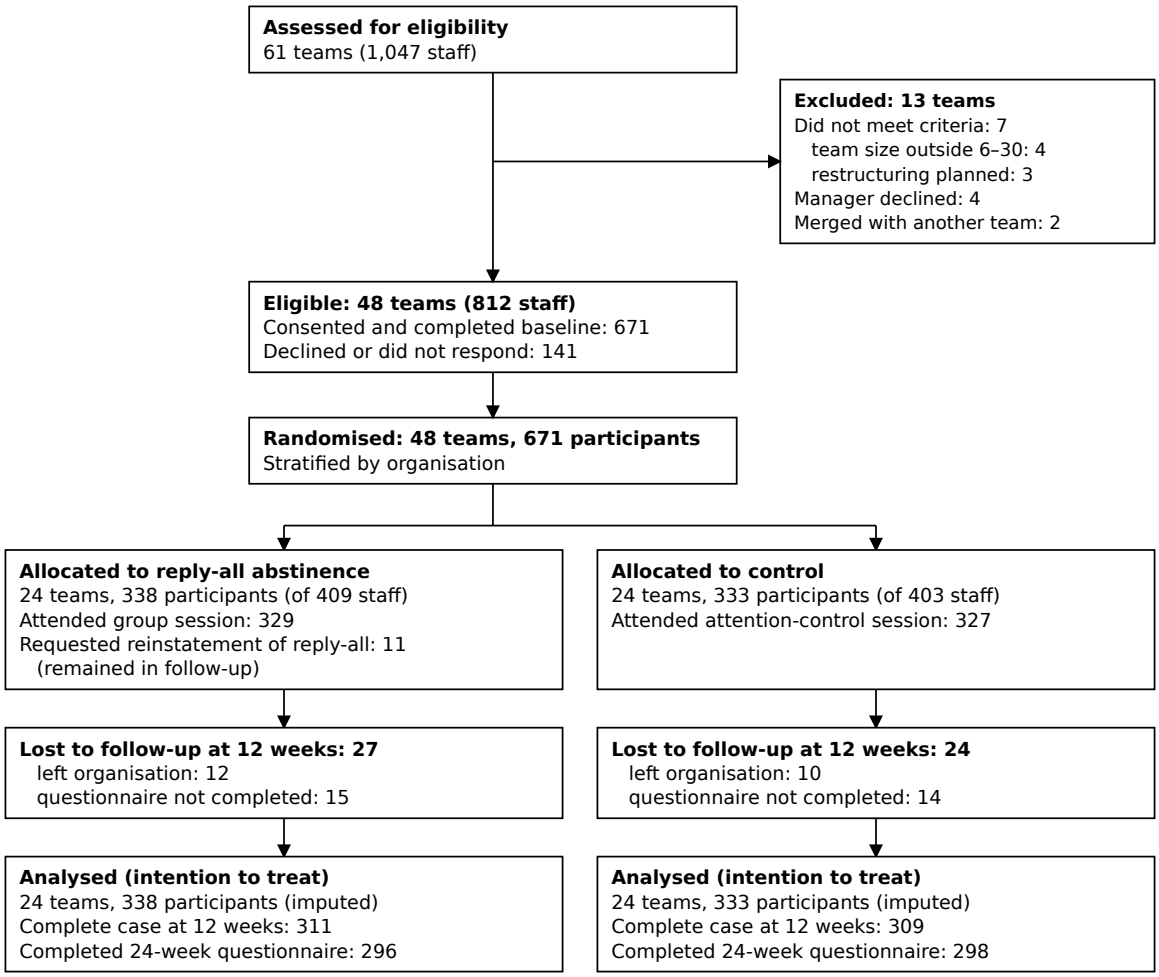


Figure 1. Trial profile. Numbers of eligible staff refer to all eligible members of the randomised teams; participants are those who consented and completed baseline assessments before allocation was disclosed.

Baseline characteristics were similar between groups (Table 1). The mean age was 41 years, a little over half of participants were women, and about a third were heavy reply-all users. Participants received a median of 60 emails per working day, of which roughly one in six was a reply-all message.

Table 1. Baseline characteristics of participants, by allocated group. Email measures are from server metadata for the 4 weeks before randomisation. CBI = Copenhagen Burnout Inventory. UWES-9 = Utrecht Work Engagement Scale.

	Reply-all abstinence (n=338)	Control (n=333)
Teams	24	24
Participants per team, median (range)	14 (7–24)	14 (6–25)
Organisation A	142 (42.0%)	139 (41.7%)
Organisation B	104 (30.8%)	101 (30.3%)
Organisation C	92 (27.2%)	93 (27.9%)
Age, years	41.3 (10.8)	40.7 (11.2)
Women	181 (53.6%)	176 (52.9%)
Full-time	287 (84.9%)	279 (83.8%)
Line-management responsibility	41 (12.1%)	38 (11.4%)
Years in organisation, median (IQR)	6 (3–12)	6 (2–11)
Works remotely at least 1 day a week	239 (70.7%)	231 (69.4%)
CBI work-related burnout (0–100)	48.6 (17.9)	47.9 (18.3)
CBI personal burnout (0–100)	45.2 (16.4)	44.8 (16.9)
UWES-9 work engagement (0–6)	3.68 (1.04)	3.71 (1.01)
Inbound emails per working day, median (IQR)	61 (38–94)	59 (36–91)
Inbound reply-all messages per working day, median (IQR)	9.4 (5.1–16.2)	9.1 (4.8–15.7)
Reply-all messages sent per week, median (IQR)	7 (3–14)	7 (3–13)
Heavy users (≥10 reply-all messages sent per week)	124 (36.7%)	119 (35.7%)

Data are n (%) or mean (SD) unless otherwise stated.

Work-related burnout decreased in the intervention group over the 12 weeks and changed little in the control group (Figure 2). At 12 weeks the mean score was 43.9 (SD 18.1) in the intervention group and 47.6 (18.6) in the control group, an adjusted mean difference of −4.1 points (95% CI −6.9 to −1.3; p=0.004). The ICC for teams was 0.041 (0.012–0.083). The complete-case estimate was −4.3 (−7.1 to −1.5), and further adjustment for age, sex and managerial role gave −4.0 (−6.8 to −1.2). Differences at 4 and 8 weeks were −1.3 (−3.9 to 1.3) and −2.8 (−5.5 to −0.1). At 24 weeks, 12 weeks after the command was restored, the difference was −1.2 (−4.1 to 1.7).

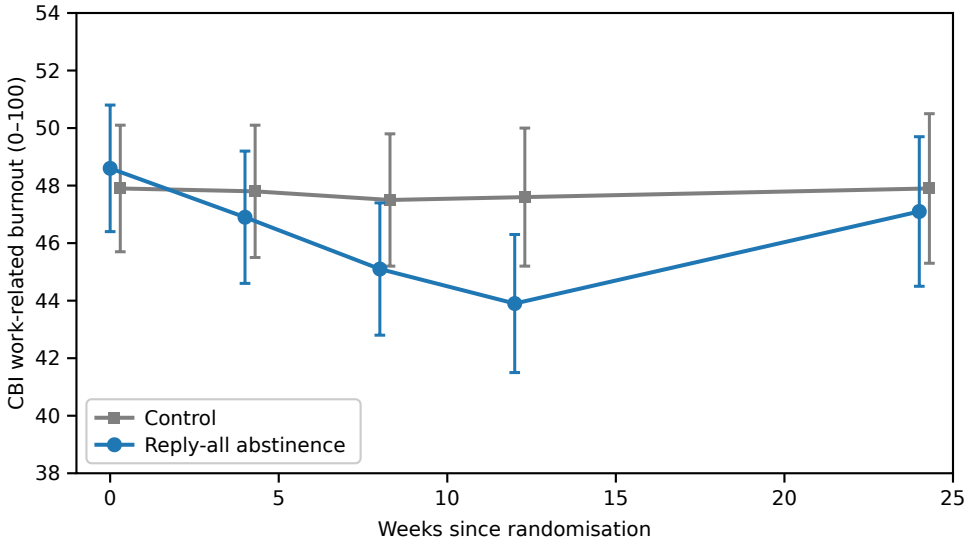


Figure 2. CBI work-related burnout by allocated group. Points are observed means with 95% confidence intervals that account for clustering. The reply-all command was re-stored in intervention teams after the 12-week assessment.

Effects on secondary outcomes are shown in Table 2. Personal burnout was lower in the intervention group, with a confidence interval that approached zero, and there was no difference in work engagement. Intervention participants received 16% fewer emails and 54% fewer reply-all messages than controls. The number of emails sent outside working hours and the number of days of sickness absence were similar in the two groups. Intervention participants were more likely to report feeling left out of the loop.

Table 2. Secondary outcomes at 12 weeks. Questionnaire outcomes are for participants who completed the 12-week assessment; email and absence outcomes are for participants still employed at 12 weeks. Effects are adjusted for baseline value and organisation and account for clustering by team. MD = mean difference. RR = rate ratio. OR = odds ratio.

Outcome	Reply-all abstinence	Control	Effect (95% CI)
CBI personal burnout, mean (SD)	41.9 (16.7), n=311	44.3 (17.1), n=309	MD −2.6 (−5.1 to −0.1)
UWES-9 work engagement, mean (SD)	3.72 (1.03), n=311	3.69 (1.05), n=309	MD 0.04 (−0.09 to 0.17)
Inbound emails per working day, median (IQR)	50 (31–78), n=326	60 (37–92), n=323	RR 0.84 (0.77 to 0.91)
Inbound reply-all messages per working day, median (IQR)	4.1 (2.0–7.6), n=326	8.9 (4.7–15.3), n=323	RR 0.46 (0.39 to 0.55)
Emails sent outside working hours per week, median (IQR)	5 (1–12), n=326	6 (2–13), n=323	RR 0.88 (0.74 to 1.04)
Sickness absence, days, mean (SD)	1.9 (3.6), n=326	2.1 (3.9), n=323	RR 0.92 (0.71 to 1.19)
Felt left out of the loop, n/N (%)	97/311 (31.2%)	57/309 (18.4%)	OR 2.03 (1.38 to 2.98)

Of the 338 intervention participants, 197 had at least one lapse during the 12 weeks, and the Kaplan–Meier estimate of continuous abstinence at day 84 was 0.42. The median time to first lapse was 50 days. Lapses came early and were concentrated among heavy users (Figure 3, panel a): 31 of 124 heavy users (25.0%) and 110 of 214 light users (51.4%) remained abstinent, and the hazard ratio for a first lapse in heavy users was 2.12 (95% CI 1.61–2.79). Of the 197 first lapses, 102 (51.8%) were replies to which the participant had added the original recipients by hand and 74 (37.6%) were sent from the mobile application before its configuration was corrected on day 18. Eleven participants asked for the command to be restored on their own account, at a median of 23 days, and each lapsed afterwards (11 first lapses, 5.6%). The remaining 10 (5.1%) were sent from shared or delegated mailboxes to which the policy did not apply.

Inbound reply-all messages fell in intervention teams within the first week, fell further after the mobile configuration was corrected, and returned to near the control level within 8 weeks of the command being restored (Figure 3, panel b). Three intervention teams asked to keep the command disabled after week 12; this was done after the 12-week assessment, and these teams are analysed as randomised at 24 weeks.

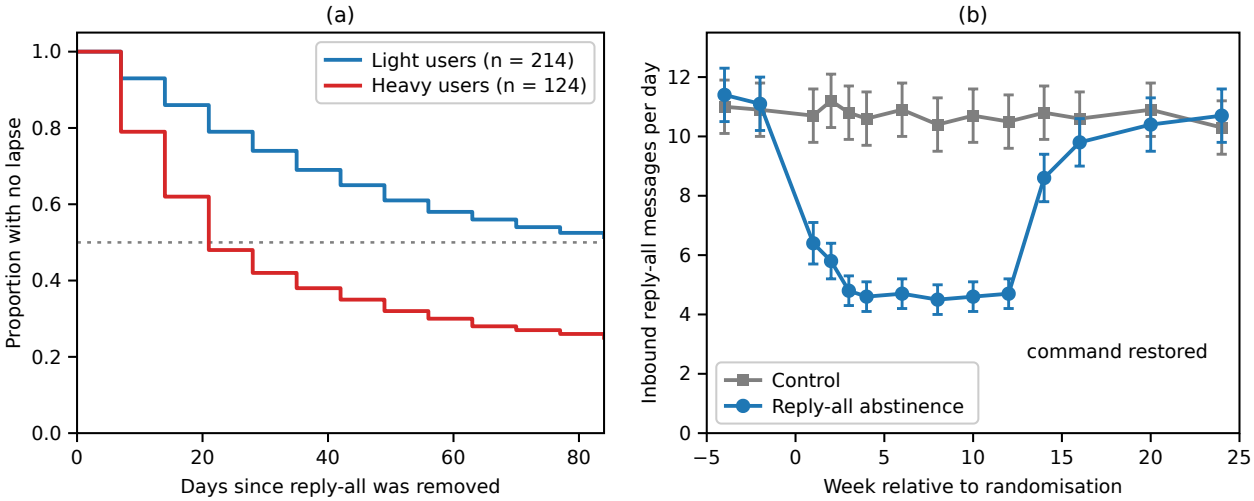


Figure 3. Abstinence and inbound reply-all messages. (a) Kaplan–Meier estimates of continuous abstinence (no lapse, as defined in equation (1)) in the intervention group, by baseline reply-all use. Heavy users sent ten or more reply-all messages per week at baseline. (b) Mean inbound reply-all messages per working day, by week, with 95% confidence intervals. The command was removed at week 0 and restored after week 12.

In pre-specified subgroup analyses the difference in work-related burnout was −5.9 (−9.8 to −2.0) in heavy users and −3.0 (−6.3 to 0.3) in light users (p for interaction 0.21), −2.2 (−8.9 to 4.5) in managers and −4.4 (−7.3 to −1.5) in other staff (p=0.53), with no evidence of heterogeneity between organisations (p=0.64). In the exploratory mediation analysis, the reduction in inbound reply-all messages accounted for an estimated 38% of the effect on work-related burnout (95% CI 12–71%).

No serious adverse events were reported. Participants reported 23 incidents of missed information in the intervention group, most often a changed meeting time or venue (9 incidents), and 11 in the control group.

Discussion

In this cluster-randomised trial, removing the reply-all command from the email clients of whole teams for 12 weeks reduced work-related burnout by about 4 points on a 100-point scale. The reduction was present at 8 weeks, was largest at 12 weeks, and had largely disappeared 12 weeks after the command was restored. Reply-all traffic into intervention teams halved, and more than half of participants lapsed at least once, most often by rebuilding the recipient list by hand.

How much the effect matters is uncertain. The estimate lies below the 5-point difference we specified as minimally important, although the confidence interval includes it. As a standardised difference the effect is about 0.23, a little smaller than the pooled effect of controlled interventions against physician burnout in one meta-analysis (standardised mean difference −0.29) [3], and the intervention costs little to deliver. Against this, almost a third of intervention participants said they had felt left out of the loop, twice the proportion of controls, and the number of reported incidents of missed information doubled. Some of the email that reply-all distributes is useful to some of the people who receive it, and the intervention removed both kinds.

The mediation analysis suggests that less than half of the effect operated through the reduction in reply-all messages received. The remainder could reflect other changes in how teams communicated, such as the conventions for group responses agreed in the

intervention session, or it could reflect expectation. Participants could not be masked and the primary outcome was self-reported. Intervention participants knew that their team was doing something intended to reduce stress. The attention-matched control session was designed to equalise contact with the research team, but it is unlikely to have equalised expectations, and research participation can itself change behaviour and reporting [22]. The objective email outcomes are not subject to this bias, and they moved in the expected direction, but they are not measures of burnout.

The pattern of lapses resembles that seen in smoking cessation trials, with most lapses occurring early and heavy users relapsing faster [10]. Half of first lapses involved deliberately reconstructing the recipient list, which takes longer than pressing a button. We did not ask participants why they did this, and a lapse by our definition is not necessarily inappropriate: some replies need to reach everyone. A definition that excluded such replies would require reading message content, which we did not do.

The trial has other limitations. The failure of the policy on one mobile application during the first 18 days weakened the intervention in two organisations, and the effect on burnout may have been larger with full delivery from the start. The cluster design protected against contamination within teams, but intervention participants continued to receive reply-all messages from people outside their team, which is reflected in the residual traffic shown in Figure 3. The three participating organisations were medium to large employers in England using the same email platform, and the 83% consent rate means that staff who declined, who may have had different email habits, are not represented. Twelve weeks is short relative to the time over which burnout develops, and we cannot say whether a longer period of abstinence would have produced a larger

or more durable effect. Three intervention teams kept the command disabled after week 12, which, if anything, would have made the 24-week difference larger.

A longer trial with outcomes that do not depend on self-report, such as absence records, would be needed to establish whether the effect on burnout is worth the cost in missed information. Recovery from work, which is thought to depend partly on psychological detachment outside working hours [23], was not measured and would be a reasonable outcome to include.

Ethics statement

The trial was approved by the Faculty of Medicine Research Ethics Committee of Calder Vale University (reference CVU-FMREC-2024-0381) and by the management of each participating organisation. All participants gave written informed consent. Email metadata were processed under data-sharing agreements with each organisation, recipient lists were hashed before transfer, and no message content was accessed.

Funding

This trial was funded by the Workplace Health Research Fund (grant WHRF-2024-112). The funder had no role in study design, data collection, data analysis, data interpretation or writing of the report.

Data sharing

De-identified individual participant data from the questionnaires, with a data dictionary, will be available from the corresponding author after publication to researchers who provide a methodologically sound proposal and sign a data access agreement. Email

metadata cannot be shared under the agreements with the participating organisations. The protocol and statistical analysis plan are available from the trial registry.

Contributors

SRA and GAP conceived the trial. SRA, MHS and PN-H designed the intervention and ran the sessions. OTB wrote the statistical analysis plan and did the statistical analysis. SRA and OTB accessed and verified the data. SRA wrote the first draft, and all authors revised it and approved the final version.

Declaration of interests

We declare no competing interests.

Acknowledgments

We thank the participating staff, the IT departments of the three organisations, which made the configuration changes, and the members of the trial steering committee.

References

1. World Health Organization. Burn-out an "occupational phenomenon": International Classification of Diseases. Departmental news, 28 May 2019; 2019.
2. Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry. 2016;15(2):103–111.
3. Panagioti M, Panagopoulou E, Bower P, Lewith G, Kontopantelis E, Chew-Graham C, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. JAMA Internal Medicine. 2017;177(2):195–205.
4. Ruotsalainen JH, Verbeek JH, Mariné A, Serra C. Preventing occupational stress in healthcare workers. Cochrane Database of Systematic Reviews. 2015(4):CD002892.
5. Barley SR, Meyerson DE, Grodal S. E-mail as a source and symbol of stress. Organization Science. 2011;22(4):887–906.
6. Kushlev K, Dunn EW. Checking email less frequently reduces stress. Computers in Human Behavior. 2015;43:220–228.

7. Mark G, Voida S, Cardello A. "A pace not dictated by electrons": an empirical study of work without email. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2012. p. 555–564.
8. Okonkwo-Reyes T, Lindahl M. Reply-all cascades in enterprise email: a metadata study of 1.4 million threads. Computers in Human Behavior. 2023;141:107612.
9. Achterberg SR, Pyke GA, Strand MH. Feasibility of removing the reply-all function in two office teams: a mixed-methods pilot study. Pilot and Feasibility Studies. 2024;10:88.
10. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicotine & Tobacco Research. 2003;5(1):13–25.
11. West R, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction. 2005;100(3):299–303.
12. Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. BMJ. 2012;345:e5661.
13. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
14. Kristensen TS, Borritz M, Villadsen E, Christensen KB. The Copenhagen Burnout Inventory: a new tool for the assessment of burnout. Work & Stress. 2005;19(3):192–207.
15. Schaufeli WB, Bakker AB, Salanova M. The measurement of work engagement with a short questionnaire: a cross-national study. Educational and Psychological Measurement. 2006;66(4):701–716.
16. Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Medical Research Methodology. 2011;11:102.
17. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics. 1997;53(3):983–997.
18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Statistics in Medicine. 2011;30(4):377–399.
19. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958;53(282):457–481.
20. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B. 1972;34(2):187–220.
21. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods. 2010;15(4):309–334.
22. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. Journal of Clinical Epidemiology. 2014;67(3):267–277.
23. Sonnentag S, Fritz C. The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work. Journal of Occupational Health Psychology. 2007;12(3):204–221.