Supplementary Materials

Postoperative Coffee and Bowel Recovery After Abdominal Surgery: An Updated Systematic Review and Meta-Analysis of Randomized Controlled Trials

Sang-ji Choi and Gibong Chae

Section S1. Search Strategies

Databases: PubMed (MEDLINE), Embase (Embase.com), and the Cochrane Central Register of Controlled Trials (CENTRAL). The searches were executed on 11 June 2026 with no language or date restriction and were updated on 11 November 2026 with the same strategies (see Update search, below). A recall-oriented strategy was used: terms for the intervention (coffee or caffeine) were combined with terms for the operative or postoperative context, without an outcome filter, so that eligible trials reporting gastrointestinal recovery in any wording were captured; outcome and design eligibility were assessed at screening. PubMed yielded 787 records.

PubMed (MEDLINE)

#1 coffee[tiab] OR caffeine[tiab] OR caffeinated[tiab] OR decaffeinated[tiab]
#2 postoperative[tiab] OR "post-operative"[tiab] OR perioperative[tiab] OR "after surgery"[tiab] OR "following surgery"[tiab] OR "after operation"[tiab] OR ERAS[tiab] OR "enhanced recovery"[tiab] OR colectomy[tiab] OR "colorectal surgery"[tiab] OR colorectal[tiab] OR "anterior resection"[tiab] OR proctectomy[tiab] OR sigmoidectomy[tiab] OR hemicolectomy[tiab] OR cesarean[tiab] OR caesarean[tiab] OR "c-section"[tiab] OR hysterectomy[tiab] OR gastrectomy[tiab] OR appendectomy[tiab] OR cholecystectomy[tiab] OR laparotomy[tiab] OR laparoscop*[tiab] OR "abdominal surgery"[tiab] OR "bowel resection"[tiab]
#3 #1 AND #2 NOT (animals[mh] NOT humans[mh]) (787 records)

Embase (Embase.com)

#1 'coffee'/exp OR coffee:ti,ab OR caffeine:ti,ab OR 'caffeine'/exp OR caffeinated:ti,ab OR decaffeinated:ti,ab
#2 'postoperative period'/exp OR postoperative:ti,ab OR perioperative:ti,ab OR 'after surgery':ti,ab OR ERAS:ti,ab OR 'enhanced recovery':ti,ab OR 'colorectal surgery'/exp OR colectomy:ti,ab OR colorectal:ti,ab OR 'anterior resection':ti,ab OR proctectomy:ti,ab OR cesarean:ti,ab OR caesarean:ti,ab OR hysterectomy:ti,ab OR gastrectomy:ti,ab OR appendectomy:ti,ab OR cholecystectomy:ti,ab OR laparotomy:ti,ab OR laparoscop*:ti,ab OR 'abdominal surgery':ti,ab
#3 #1 AND #2 NOT [animals]/lim

Cochrane CENTRAL (Search Manager)

#1 (coffee OR caffeine OR caffeinated OR decaffeinated):ti,ab,kw

#2 (postoperative OR perioperative OR "enhanced recovery" OR colectomy OR "colorectal surgery" OR colorectal OR "anterior resection" OR proctectomy OR cesarean OR caesarean OR hysterectomy OR gastrectomy OR appendectomy OR cholecystectomy OR laparotomy OR laparoscop* OR "abdominal surgery"):ti,ab,kw
#3 #1 AND #2

Update search (11 November 2026)

The registered strategies above were rerun unchanged on 11 November 2026, and only records added after the original search were retained.
PubMed: strategy #3 returned 818 records; the 32 PMIDs absent from the original export of 787 records were retrieved (an entry-date limit of 11 June to 11 November 2026 returned 31 of them).
Embase: strategy #3 (1886 records on 11 November 2026) was combined with [11-6-2026]/sd, giving 84 records (74 articles and 10 clinical-trial registry records).
CENTRAL: a publication-date limit returned no records, so the Trials result of strategy #3 (328 records) was compared with the original export (319 records) by CENTRAL identifier, giving 9 new records.
In total, 125 records were retrieved; 30 were duplicates within the update and 13 were already among the original records, leaving 82 records for title and abstract screening. One full-text report was assessed and excluded (Table S3).

Section S2. Citation Tracking and Update-Search Screening

Forward and backward citation tracking was performed on 11 November 2026 for 29 seed records: the 18 trials included at submission, Simsek 2022, and 10 prior syntheses (Watanabe 2021, Vaghiri 2024, Yao 2026, Sinz 2023, a 2026 meta-analysis of coffee after cesarean section, Eamudomkarn 2018, Gkegkes 2020, Kane 2020, Yang 2022, and Cornwall 2020). Referenced and citing works were retrieved from OpenAlex (one seed, Ghomian 2022, could not be resolved there) and supplemented by the reference lists of the available full-text reports. Of 704 linked works, 175 mentioned coffee, caffeine, espresso, or decaffeinated coffee in the title or abstract; 98 of these were not already among the seed records or the original or updated database records and were screened by title and abstract, and five were assessed in full text. One further trial cited in peer review was located by web searching and assessed in full text.
The two eligible trials (Simsek 2022 and Jabeen 2024) were published in journals not indexed in PubMed, Embase, or CENTRAL, and both were cited by an included trial (Tuzmen 2025); they had not been identified at submission. Neither reported trial registration. Simsek 2022 gave different arm sizes in its text (30 and 35) and in its table and flow diagram (31 and 34; used here). Jabeen 2024 did not describe how the random sequence was generated or concealed, randomized 60 participants per arm but analyzed 56 in the coffee arm without explanation, reported standard deviations of 5.6 to 7.8 hours for mean hospital stays of 72 and 101 hours, and printed one table value that disagrees with its text (a row not used in this review). These features informed the RoB 2 judgments and the post hoc sensitivity analysis that excludes both trials (Table S2). Full-text decisions for all reports assessed after submission are listed in Table S3.

Section S3. Computational Implementation and Verification

Software. All analyses used Python 3.10 with the standard library only (math and statistics.NormalDist). The scripts, the extracted dataset, and the cached outputs are deposited with this article; running them in the order given in the accompanying README reproduces every estimate reported in the manuscript and in this supplement.

Effect sizes and conversions. For each trial, the mean difference (MD) and its standard error were computed from arm-level means and standard deviations (SD). Medians with interquartile ranges were converted with the estimators of Wan et al. (mean = (q1 + median + q3)/3; SD = (q3 − q1)/$\eta(n)$, with $\eta(n) = 2\Phi^{-1}[(0.75n - 0.125)/(n + 0.25)]$); medians with minimum and maximum were converted with the corresponding estimators (mean = (min + 2 × median + max)/4; SD = (max − min)/$\xi(n)$, with $\xi(n) = 2\Phi^{-1}[(n - 0.375)/(n + 0.25)]$); for the one trial that reported medians with 95% confidence intervals, the median was used as the mean and SD = $\sqrt{n} \times$ (upper − lower)/(2 × $t_{0.975, n-1}$). Times were harmonized to hours (length of stay to days). Dichotomous outcomes used the log risk ratio, with 0.5 added to all cells of trials with a single zero cell and exclusion of double-zero trials.

Between-trial variance. τ^2 was estimated by restricted maximum likelihood with the fixed-point iteration $\tau^2_{new} = \max\{0, \Sigma w_i^2[(y_i - \hat{\mu})^2 - v_i]/\Sigma w_i^2 + 1/\Sigma w_i\}$, where $w_i = 1/(v_i + \tau^2)$, started from the DerSimonian–Laird value and stopped when successive values differed by less than 10^{-10} (at most 500 iterations). Paule–Mandel estimation (bisection on the generalized Q statistic) was prespecified as a fallback if the iteration did not converge; in Revision 1 all 109 REML estimations converged (within 80 iterations), so the fallback was never used.

Confidence and prediction intervals. The pooled estimate $\hat{\mu} = \Sigma w_i y_i/\Sigma w_i$ was accompanied by the Hartung–Knapp–Sidik–Jonkman variance $q/\Sigma w_i$, with $q = \Sigma w_i(y_i - \hat{\mu})^2/(k - 1)$ and no truncation of q at 1; the 95% confidence interval and p value used a t distribution with k − 1 degrees of freedom. The 95% prediction interval was $\hat{\mu} \pm t_{0.975, k-2} \times \sqrt{(\tau^2 + 1/\Sigma w_i)}$ for $k \geq 3$. I^2 was derived from Cochran's Q with inverse-variance weights. Subgroup differences were tested with a between-subgroup Q statistic on the subgroup REML estimates (weights $1/SE^2$, χ^2 with g − 1 degrees of freedom), restricted to subgroups with at least three trials. Small-study effects were assessed with the Egger regression of y_i/SE_i on $1/SE_i$ (t test of the intercept, k − 2 degrees of freedom) for outcomes with at least ten trials.

Verification. (i) Running the deposited code on the dataset of the original submission reproduces the submitted outputs exactly. (ii) An independent implementation that shares no code with the analysis engine, in which τ^2 is obtained by direct maximization of the restricted likelihood and the t distribution by numerical integration (verify_independent.py), reproduced all 65 pooled analyses (primary and leave-one-out, continuous and binary) with a largest absolute difference of 1.0×10^{-5} in any estimate, interval limit, τ^2, I^2, or p value. (iii) Values for the two added trials were recomputed by hand from the tables of their reports.

Supplementary Tables

Table S1. Outcome-specific risk-of-bias (RoB 2) judgments for all outcomes.

Study	Outcome	D1	D2	D3	D4	D5	Overall
Bozkurt Koseoglu 2020	Stool	Low	Some concerns	Low	Some concerns	Low	Some concerns
Dulskas 2015	Stool	Some concerns	High	Some concerns	Some concerns	Some concerns	High
Ghomian 2022	Stool	Low	Low	Some concerns	Some concerns	Some concerns	Some concerns
Gungorduk 2017	Stool	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Gungorduk 2020	Stool	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Hasler-Gehrer 2019	Stool	Low	Some concerns	Low	Some concerns	Low	Some concerns

Holland 2023	Stool	Low	High	Some concerns	High	Low	High
Jabeen 2024	Stool	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Kao 2025	Stool	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Muller 2012	Stool	Low	Some concerns	Low	Some concerns	Low	Some concerns
Nasseri 2023	Stool	Low	High	Low	Some concerns	Low	High
Rabiepoor 2018	Stool	Some concerns	High	Low	High	Some concerns	High
Songthamwat 2024	Stool	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Tuzmen 2025	Stool	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Zamanabadi 2021	Stool	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Bozkurt Koseoglu 2020	Flatus	Low	Some concerns	Low	Some concerns	Low	Some concerns
Ghomian 2022	Flatus	Low	Low	Some concerns	Some concerns	Some concerns	Some concerns
Gungorduk 2017	Flatus	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Gungorduk 2020	Flatus	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Hasler-Gehrer 2019	Flatus	Low	Some concerns	Low	Some concerns	Low	Some concerns
Holland 2023	Flatus	Low	High	Some concerns	High	Low	High
Jabeen 2024	Flatus	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Kao 2025	Flatus	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Muller 2012	Flatus	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Nasseri 2023	Flatus	Low	High	Low	Some concerns	Low	High
Rabiepoor 2018	Flatus	Some concerns	High	Low	High	Some concerns	High
Simsek 2022	Flatus	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Songthamwat 2024	Flatus	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Zamanabadi 2021	Flatus	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Ghomian 2022	POI	Low	Low	Some concerns	Some concerns	Some concerns	Some concerns
Gungorduk 2017	POI	Low	Some concerns	Low	Low	Some concerns	Some concerns
Gungorduk 2020	POI	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Songthamwat 2024	POI	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns

Bozkurt Koseoglu 2020	Solid Diet	Low	Some concerns	Low	Some concerns	Low	Some concerns
Gungorduk 2017	Solid Diet	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Gungorduk 2020	Solid Diet	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Muller 2012	Solid Diet	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Songthamwat 2024	Solid Diet	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Ghomian 2022	LOS	Low	Low	Some concerns	Low	Some concerns	Some concerns
Gungorduk 2017	LOS	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Gungorduk 2020	LOS	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Hasler-Gehrer 2019	LOS	Low	Some concerns	Low	Some concerns	Low	Some concerns
Holland 2023	LOS	Low	Some concerns	Some concerns	Some concerns	Low	Some concerns
Jabeen 2024	LOS	Some concerns	Some concerns	Low	Some concerns	Some concerns	Some concerns
Muller 2012	LOS	Low	Some concerns	Low	Some concerns	Low	Some concerns
Nasseri 2023	LOS	Low	Some concerns	Low	Low	Low	Some concerns
Rabiepoor 2018	LOS	Some concerns	High	Low	Some concerns	Some concerns	High
Songthamwat 2024	LOS	Low	Some concerns	Low	Low	Some concerns	Some concerns
Gungorduk 2017	PONV	Low	Some concerns	Low	Some concerns	Some concerns	Some concerns
Hayashi 2019	PONV	Low	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Nasseri 2023	PONV	Low	Some concerns	Low	Some concerns	Low	Some concerns
Rabiepoor 2018	PONV	Some concerns	High	Low	High	Some concerns	High
Dulskas 2015	AE	Some concerns	High	Some concerns	Some concerns	Some concerns	High
Hasler-Gehrer 2019	AE	Low	Low	Low	Low	Low	Low
Hayashi 2019	AE	Low	Some concerns	Some concerns	Some concerns	Some concerns	Some concerns
Holland 2023	AE	Low	Some concerns	Some concerns	Low	Some concerns	Some concerns
Muller 2012	AE	Low	Some concerns	Low	Some concerns	Low	Some concerns
Nasseri 2023	AE	Low	Some concerns	Low	Some concerns	Low	Some concerns

Each row is one trial-by-outcome assessment. D1, randomization process; D2, deviations from intended interventions; D3, missing outcome data; D4, measurement of the outcome; D5, selection of the reported result. Overall judgment follows the standard RoB 2 algorithm (any High → High; otherwise any Some concerns → Some concerns; otherwise Low).

Note. Figure 4 presents the traffic-light and summary plots for the primary outcome (time to first stool). This table provides the corresponding outcome-specific judgments underpinning every GRADE rating.

Table S2. Post hoc sensitivity analyses for the outcomes with ten or more trials (REML with the Hartung–Knapp–Sidik–Jonkman adjustment).

Outcome	Analysis	Trials	Participants	MD (95% CI)	I^2 (%)	p
Time to first stool (hours)	Primary analysis (caffeinated coffee)	15	1291	−6.93 (−9.81 to −4.05)	81	<0.001
Time to first stool (hours)	(A) Trials reporting means with SD only	13	1080	−6.81 (−10.18 to −3.44)	83	<0.001
Time to first stool (hours)	(B) Volume-matched beverage comparator	11	901	−4.72 (−6.93 to −2.50)	25	<0.001
Time to first stool (hours)	(B) Usual care or no intervention	4	390	−9.84 (−19.07 to −0.62)	95	0.043
Time to first stool (hours)	(C) Excluding Simsek 2022 and Jabeen 2024	14	1175	−6.59 (−9.71 to −3.48)	77	<0.001
Time to first stool (hours)	Prespecified decaffeinated subgroup	3	172	−11.41 (−42.18 to 19.36)	78	0.252
Time to first flatus (hours)	Primary analysis (caffeinated coffee)	14	1239	−3.19 (−6.51 to 0.13)	93	0.059
Time to first flatus (hours)	(A) Trials reporting means with SD only	10	920	−5.00 (−7.70 to −2.30)	75	0.002
Time to first flatus (hours)	(B) Volume-matched beverage comparator	11	906	−1.88 (−5.62 to 1.85)	92	0.288
Time to first flatus (hours)	(B) Usual care or no intervention	3	333	−7.11 (−18.57 to 4.35)	91	0.116
Time to first flatus (hours)	(C) Excluding Simsek 2022 and Jabeen 2024	12	1058	−2.66 (−6.63 to 1.31)	94	0.168
Length of stay (days)	Primary analysis (caffeinated coffee)	10	961	−0.57 (−1.12 to −0.02)	95	0.044
Length of stay (days)	(A) Trials reporting means with SD only	8	778	−0.70 (−1.40 to 0.01)	95	0.052
Length of stay (days)	(B) Volume-matched beverage comparator	8	731	−0.28 (−0.69 to 0.13)	67	0.154
Length of stay (days)	(B) Usual care or no intervention	2	230	not pooled (<3 trials)		

Length of stay (days)	(C) Excluding Simsek 2022 and Jabeen 2024	9	845	−0.41 (−0.94 to 0.11)	71	0.108

MD, mean difference (coffee minus control; negative values favor coffee); CI, confidence interval; SD, standard deviation. (A) excludes trials whose means and SDs were estimated from medians; (B) splits trials by comparator; (C) excludes the two trials identified by citation tracking at revision, both published in journals not indexed in the searched databases. The decaffeinated subgroup compares decaffeinated arms with their control arms and is shown for the only outcome with at least three such trials.

Table S3. Reports assessed in full text after the original submission (update search, citation tracking, and web searching).

Report	Source	Decision	Reason
Besen et al., J. Obstet. Gynaecol. Res. 2026, 52, e70436 (doi:10.1111/jog.70436)	Update search	Excluded	No extractable gastrointestinal-recovery outcome: 24 h follow-up; time to flatus reported only for women with flatus within 12 h; no time to defecation
Duerust Sakalli et al., Mediterr. Nurs. Midwifery 2026 (doi:10.65717/MNM.2026.26649)	Web search	Excluded	No extractable gastrointestinal-recovery outcome: single drink at 6 h, outcomes recorded as present or absent up to 6 h
Simsek et al., Med. Sci. Int. Med. J. 2022, 11, 651–655	Citation tracking	Included	RCT; cesarean delivery; Turkish coffee versus warm water; time to first flatus
Jabeen et al., Ann. Abbasi Shaheed Hosp. Karachi Med. Dent. Coll. 2024, 29, 152–158	Citation tracking	Included	RCT; major gynecologic surgery; coffee versus no coffee; time to first flatus, time to first stool, length of stay
Mohamed et al., Al-Azhar Int. Med. J. 2023, 4, 253–256 (doi:10.58675/2682-339X.2008)	Citation tracking	Excluded	Non-randomized (prospective cohort)
Lietuvos Chirurgija 2014 (doi:10.15388/lietchirur.2014.3084)	Citation tracking	Excluded	Interim report of an included trial (Dulskas 2015)
Rev. Colomb. Cir. 2024 (doi:10.30944/20117582.2524)	Citation tracking	Excluded	Non-randomized (convenience sampling)

RCT, randomized controlled trial. The update search covered records added to PubMed, Embase, and CENTRAL between 11 June and 11 November 2026 (Appendix S1); citation tracking is described in Appendix S2.

Supplementary Figures

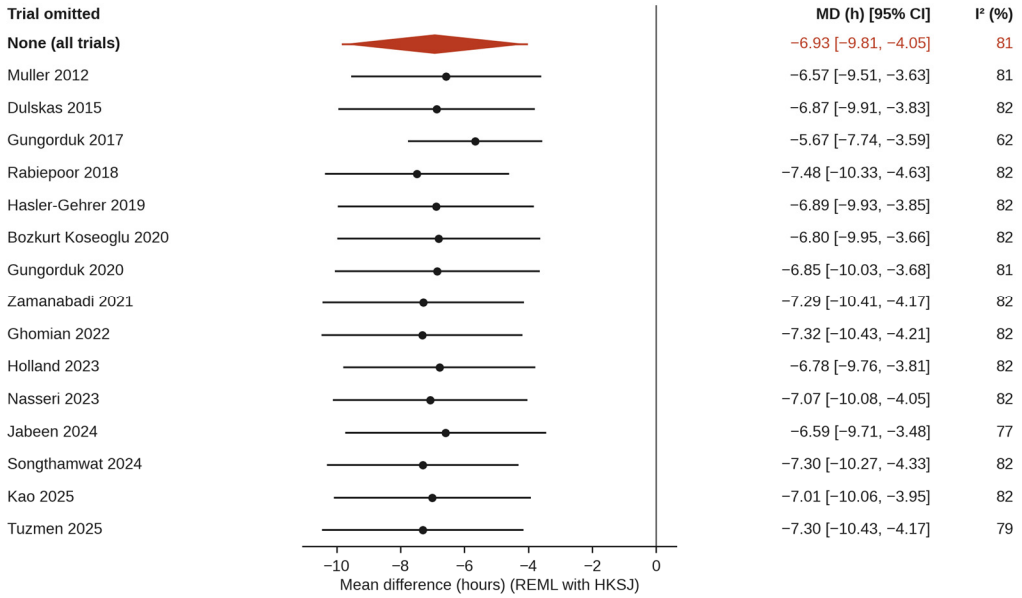


Figure S1. Leave-one-out sensitivity analysis for time to first stool. Each estimate is the pooled mean difference (hours; coffee versus control) recalculated after omitting the indicated trial (random-effects REML with the Hartung–Knapp–Sidik–Jonkman adjustment). Negative values favor coffee; the first row shows the estimate with all 15 trials.

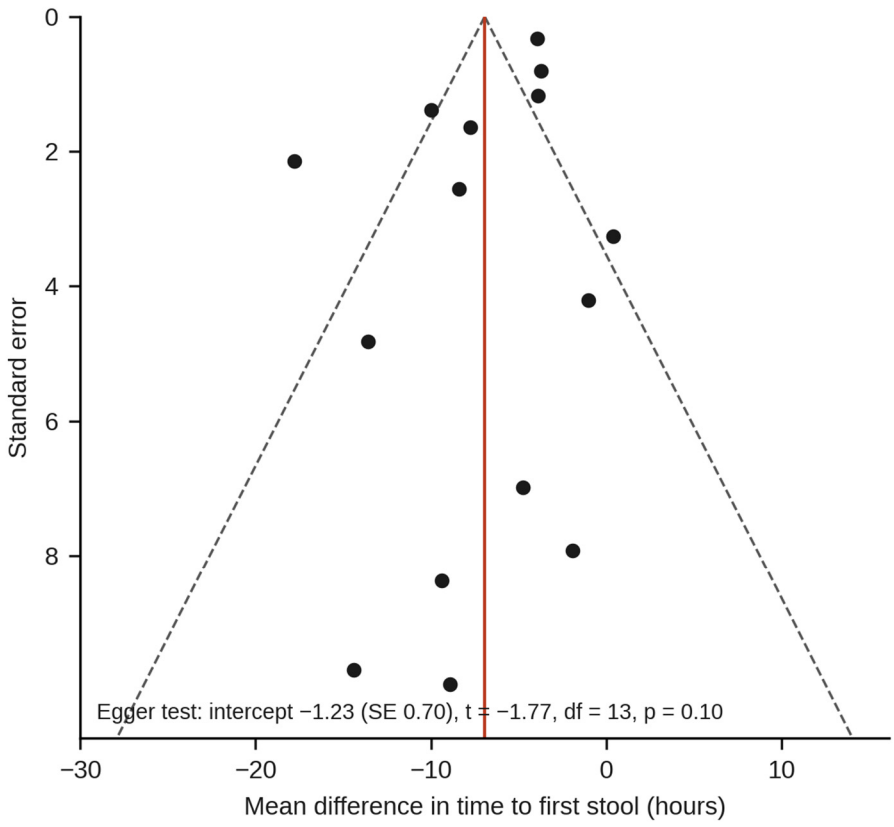


Figure S2. Funnel plot for time to first stool, used to assess small-study effects; the Egger test p value is reported in the main text.

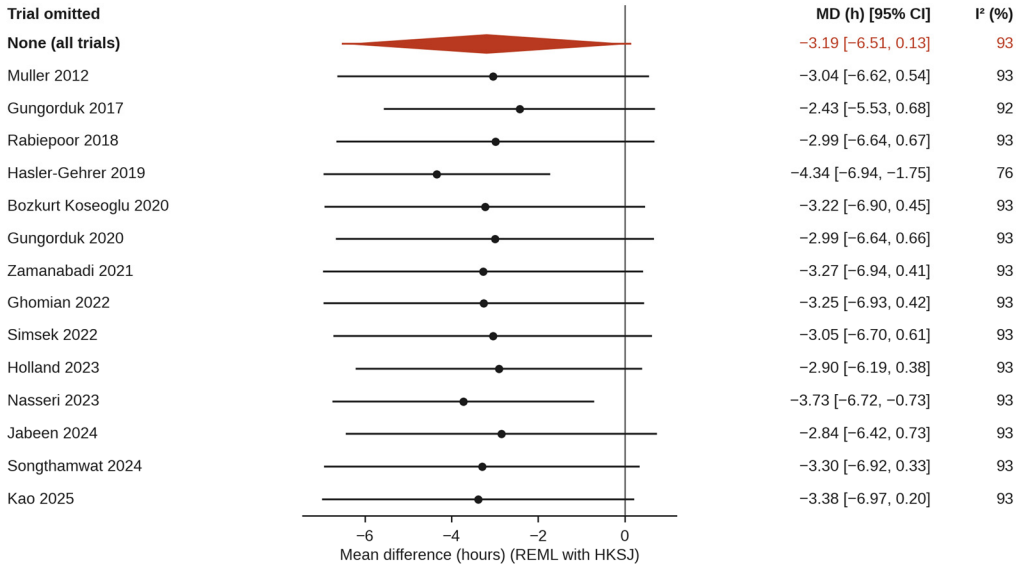


Figure S3. Leave-one-out sensitivity analysis for time to first flatus (mean difference, hours). Negative values favor coffee.

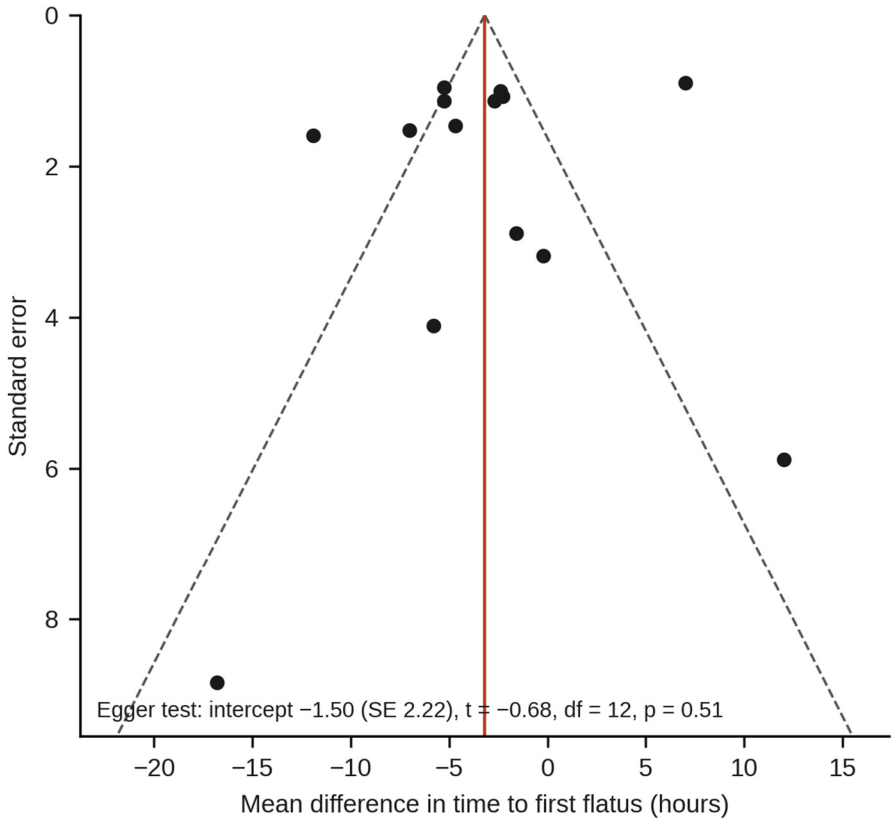
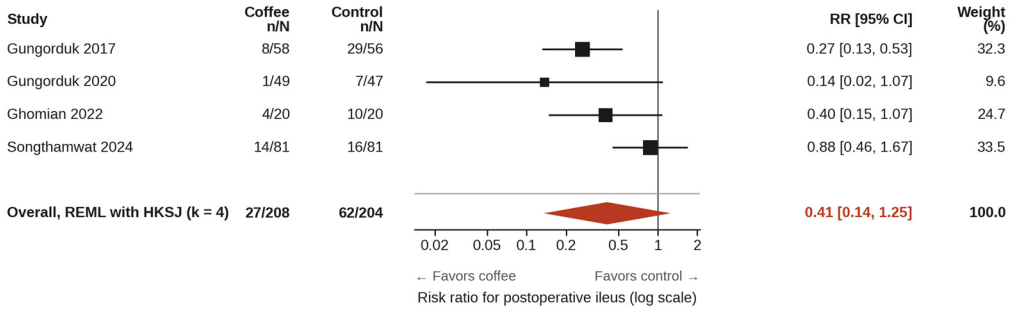
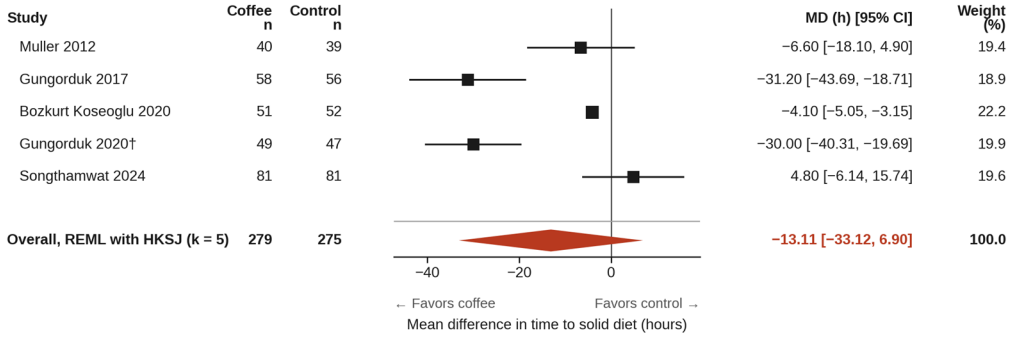


Figure S4. Funnel plot for time to first flatus.



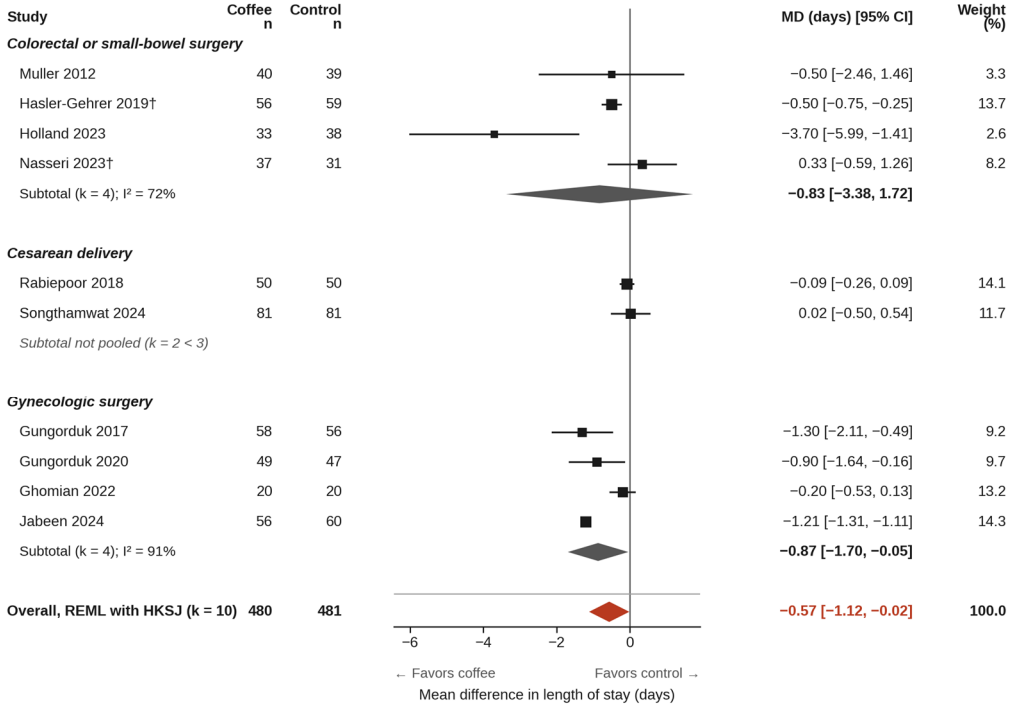
Heterogeneity: $\tau^2 = 0.29$; $I^2 = 60\%$; $Q = 7.6$, $df = 3$ ($p = 0.055$). Overall effect: $t = -2.54$, $df = 3$ ($p = 0.085$).
Mantel–Haenszel RR without continuity correction (sensitivity): 0.43 [0.29, 0.64].

Figure S5. Forest plot of postoperative ileus (risk ratio, coffee versus control; random-effects model). Values below 1 favor coffee.



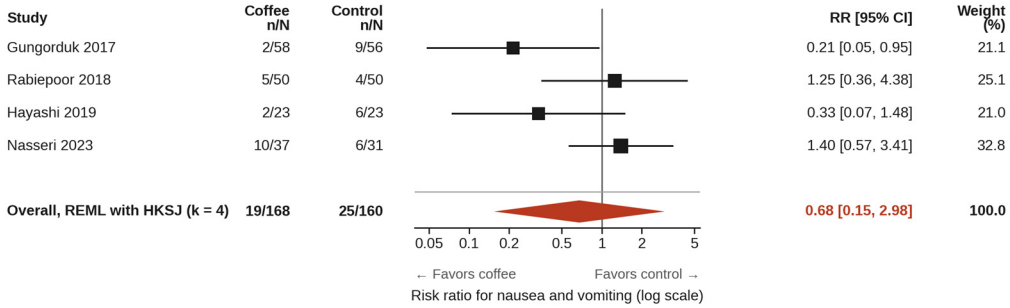
Heterogeneity: $\tau^2 = 229.89$; $I^2 = 91\%$; $Q = 44.6$, $df = 4$ ($p < 0.001$). Overall effect: $t = -1.82$, $df = 4$ ($p = 0.143$).
95% prediction interval: −66.47 to 40.24 h.
† Mean and SD estimated from median-based summaries (see Methods).

Figure S6. Forest plot of time to tolerance of solid diet (mean difference, hours). Negative values favor coffee.



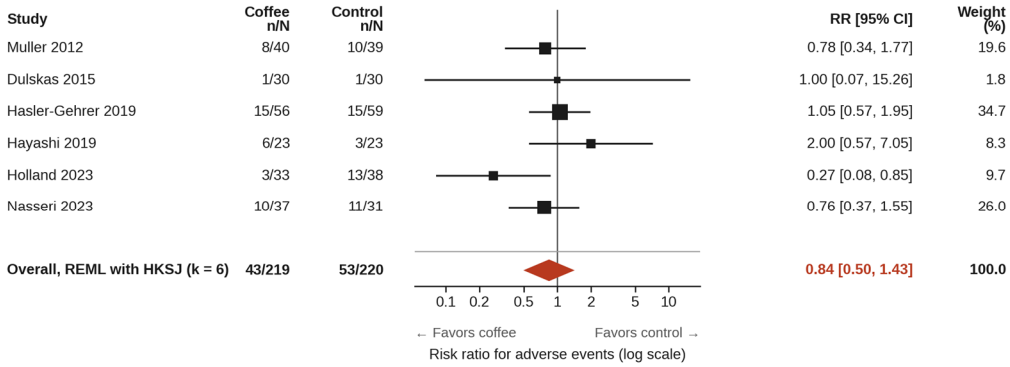
Heterogeneity: $\tau^2 = 0.30$; $I^2 = 95\%$; $Q = 165.8$, $df = 9$ ($p < 0.001$). Overall effect: $t = -2.35$, $df = 9$ ($p = 0.044$).
95% prediction interval: −1.91 to 0.77 days.
Test for subgroup differences (subgroups with ≥ 3 trials): $Q = 0.00$, $df = 1$ ($p = 0.959$).
† Mean and SD estimated from median-based summaries (see Methods).

Figure S7. Forest plot of length of hospital stay (mean difference, days), grouped by surgery type. Negative values favor coffee.



Heterogeneity: $\tau^2 = 0.45$; $I^2 = 52\%$; $Q = 6.3$, $df = 3$ ($p = 0.097$). Overall effect: $t = -0.84$, $df = 3$ ($p = 0.464$).
Mantel–Haenszel RR without continuity correction (sensitivity): 0.70 [0.40, 1.23].

Figure S8. Forest plot of postoperative nausea and vomiting (risk ratio). Values below 1 favor coffee.



Heterogeneity: $\tau^2 = 0.00$; $I^2 = 19\%$; $Q = 6.2$, $df = 5$ ($p = 0.286$). Overall effect: $t = -0.84$, $df = 5$ ($p = 0.442$).
Mantel–Haenszel RR without continuity correction (sensitivity): 0.81 [0.57, 1.15].

Figure S9. Forest plot of adverse events (risk ratio). Values below 1 favor coffee.

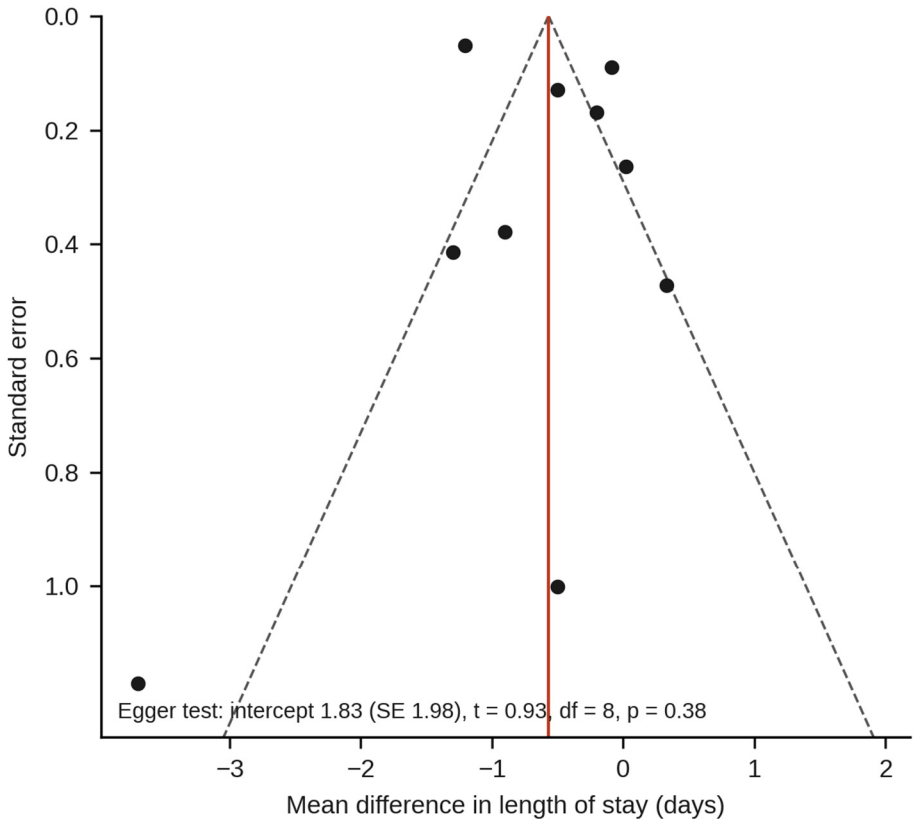


Figure S10. Funnel plot for length of hospital stay (mean difference, days); the Egger test is shown in the panel.

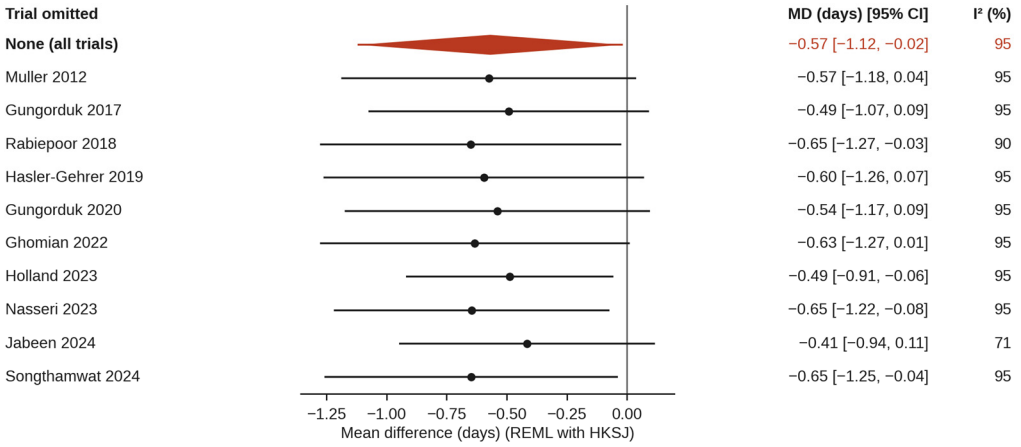


Figure S11. Leave-one-out sensitivity analysis for length of hospital stay (mean difference, days; REML with the Hartung–Knapp–Sidik–Jonkman adjustment). The first row shows the estimate with all ten trials.