Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Gut microbiota and suicidal behaviour: a phenotype-stratified systematic review and a 131-genus two-sample Mendelian randomization study

Songqing Liu[1]
Guangzhou Psychiatric Hospital
Jin Zhao[1]
Guangzhou Psychiatric Hospital
Zezhi Li[1]
Guangzhou Psychiatric Hospital
Xiaobing Lu
136888763460139.com

Guangzhou Medical University

Research Article

Keywords: gut microbiota, suicidal ideation, suicide attempt, systematic review, Mendelian randomization, evidence triangulation, alpha diversity, Flavonifractor

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11643934/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background.

Observational studies link gut microbiota to suicidal behaviour, but confounding and reverse causation limit causal inference. Mendelian randomization (MR) can address these biases, yet the two evidence streams estimate different quantities.

Methods.

Part A was a systematic review of human intestinal bacterial profiling studies with separable suicidal phenotypes, stratified by phenotype and comparator and synthesised with SWiM (PROSPERO CRD420261543928). Part B screened 131 genera with two-sample MR using MiBioGen instruments and a European-ancestry suicide-attempt GWAS (31,094 cases, 758,690 controls), with Benjamini–Hochberg control, pre-specified sensitivity, directionality and exposure-side replication. One included cohort supplied sample-level Shannon data, which were used for post hoc digitization validation and a mixed exact/digitized sensitivity analysis.

Results.

Part A included 13 reports (11 cohorts); none provided group-level Shannon or Chao1 statistics. Digitized data from three cohorts gave a pooled Shannon difference of − 0.11 (95% CI − 0.96 to 0.73); with exact data for one cohort, the post hoc estimate was − 0.16 (− 0.77 to 0.46). Risk of bias was high or critical in 11 of 13 reports. In Part B, *Flavonifractor* alone reached FDR significance (OR 0.819, q = 0.026), was classified as suggestive, and was not replicated at the exposure side (OR 1.095; R1-C). No genus had both replicated observational and robust genetic evidence.

Conclusions.

The two designs converge on the absence of robust evidence. This is an evidence boundary rather than a causal or clinical finding.

Background

Suicidal thoughts and behaviours are common in affective disorders and remain leading contributors to premature death worldwide [1]. The phenotype spans suicidal ideation (SI), suicide attempt (SA) and suicide death, which are related but not interchangeable [2]. Clinical prediction is modest [3], suicide attempt has a substantial genetic component [4], and modifiable exposures that could inform prevention remain scarce.

The gut microbiota communicates with the central nervous system through immune, metabolic, vagal and neuroendocrine pathways [5, 6]. In humans, microbial composition has been associated with depression [7], and cross-diagnostic meta-analyses report altered diversity in psychiatric disorders [8]. Direct studies of suicidal behaviour report differences in diversity or composition across ideation, attempt and death [9, 10, 11, 12, 13, 14, 15, 16], whereas a large psychiatric inpatient cohort found no association [17]. A systematic review concluded that associations were suggestive but inconclusive [18], and later narrative reviews reached similarly cautious positions [19, 20]; we are not aware of a published meta-analysis of gut bacterial diversity in suicidal phenotypes.

Observational microbiome studies have structural limits for causal inference. Diagnosis, severity, medication, diet and sampling conditions may confound both group membership and microbial measurements, and sampling after an attempt or death introduces temporal and selection problems [5, 8]. Comparators differ in meaning, phenotype ascertainment varies from dedicated instruments to single depression items [21], and reporting is often incomplete: diversity may appear only as box plots and only selected taxa may be reported [22, 23].

MR uses genetic variants strongly associated with an exposure as instruments and is less susceptible to confounding and reverse causation when the instrumental-variable assumptions hold [24, 25]. It has been applied to the gut microbiome since large-scale microbial GWAS became available [26, 27], but microbiome MR faces specific difficulties: genetic associations with taxa replicate poorly [28], measurement is platform dependent, and power is limited [29]. Previous MR studies examined major depressive disorder [30], a 15-disorder panel including suicide attempt [31], non-suicidal self-injury [32] and depressive symptoms [33].

The two evidence streams are usually presented as mutually confirmatory. They are not: observational studies estimate associations between measured microbial composition and suicidal phenotypes in particular populations, whereas MR estimates the effect of genetically predicted abundance on a GWAS-defined outcome in a different ancestry. The estimands, exposures, outcomes and scales differ, so the two cannot be pooled or treated as estimates of one causal parameter. They can still be complementary, and comparing their directions and gaps, with explicit acknowledgment of what each design cannot show, is a defensible narrative triangulation rather than formal triangulation.

We therefore conducted a two-part study: a phenotype- and comparator-stratified systematic review of human observational studies (Part A), and a two-sample MR screen of 131 bacterial genera for suicide attempt with pre-specified sensitivity, directionality and exposure-side replication (Part B). We then examined, descriptively, where the two parts agree, diverge and lack a counterpart (Table 1; Additional file 1: Figure S1).

Table 1
Design and data sources of the two study parts.

Feature	Part A: observational systematic review	Part B: two-sample Mendelian randomization
Question	Are intestinal microbiota features associated with SI, SA or suicide death, and how much can be pooled?	Is genetically predicted genus abundance causally associated with suicide attempt?
Design	Systematic review with structured synthesis; exploratory digitized sensitivity analysis	Two-sample MR with pre-specified sensitivity, directionality and exposure-side replication
Data	13 reports from 11 cohorts; 16S rRNA and shotgun metagenomics; sample-level Shannon data for one cohort	MiBioGen (18,340 participants, 24 cohorts; 131 genera); suicide-attempt GWAS [4] EUR noutah (31,094 cases / 758,690 controls); Nordic replication data
Phenotype	SI (direct, item, continuous), SA, suicide death, proxy/mixed	Suicide attempt (GWAS-defined)
Comparator	Phenotype-aligned psychiatric controls; healthy controls	Genetically predicted abundance (instrumental-variable design)
Main estimate	Hedges' g (only digitized sensitivity; no primary estimate)	OR per 1 SD genetically predicted abundance
Registration	PROSPERO CRD420261543928	Internally frozen statistical analysis plan (SAP); replication plan internally pre-registered

CS, cross-sectional; CC, case–control; CO, cohort; EUR noutah, European-ancestry suicide-attempt GWAS excluding the Utah cluster; GWAS, genome-wide association study; MR, Mendelian randomization; OR, odds ratio; SAP, statistical analysis plan; SD, standard deviation; SI, suicidal ideation; SA, suicide attempt.

Methods

Part A: systematic review

Protocol and reporting. The protocol was registered in PROSPERO (CRD420261543928) and the review is reported according to PRISMA 2020 and PRISMA-S, with SWiM principles for structured synthesis [34–36]; see Additional file 2. The eligibility scope was restricted to English-language reports; methods were clarified as described below.

Eligibility and searches. Human observational studies were eligible if they profiled intestinal bacteria comprehensively by 16S rRNA gene sequencing or shotgun metagenomics and reported a separable suicidal phenotype. Living-participant stool was the principal specimen; swabs and post-mortem content were supplementary strata. Non-intestinal microbiota, culture-only or targeted assays, animal studies and reviews were excluded. Searches ran from 19 to 29 October 2026 in PubMed/MEDLINE, Europe PMC, Embase, APA PsycInfo, CINAHL, Web of Science Core Collection, Scopus and CENTRAL,

supplemented by Google Scholar. The eight sources yielded 1,221 database and register records; 696 duplicates were removed, leaving 525 records for screening. Google Scholar contributed 104 records, which were screened separately, giving 1,325 source records in total; full strategies are in Additional file 1.

Selection and extraction. Two reviewers (Songqing Liu and Jin Zhao) independently screened titles/abstracts, assessed full reports, extracted data and appraised study quality; disagreements were resolved by consensus or by a third reviewer (Xiaobing Lu). After deduplication of the database and register records, 525 records were screened, 45 reports assessed, 32 excluded and 13 reports included. Extraction covered design, cohort links, participant characteristics, phenotype wording and window, comparator, medication and lifestyle, specimen and sequencing, and every reported result with its analysed group sizes, scale, variance, adjustment set and source location. Requests for missing group-level statistics were sent to corresponding authors after the analysis cutoff; one team [10] provided the de-identified sample-level data described below.

Phenotype and comparator framework. Phenotypes were classified into six non-exchangeable layers: SI-direct (dedicated instrument), SI-item (single depression item with an SI threshold), continuous SI (severity score), SA, suicide death and proxy/mixed. Comparators were phenotype-aligned psychiatric controls, lifetime-clean psychiatric controls, unclear psychiatric controls, healthy or general-population controls. SI and SA, psychiatric and healthy comparisons, current and lifetime measures, and different instruments were never combined.

Outcomes and synthesis. Shannon diversity and Chao1 richness were the principal outcomes. A synthesis unit was defined by phenotype layer and window, measurement tier, comparator, specimen, metric or taxon, scale and temporal design. Quantitative synthesis required at least three independent cohorts with clinically comparable estimates and valid variances, pooled as Hedges' g with random-effects restricted maximum likelihood and the modified Knapp–Hartung adjustment [37]; heterogeneity was summarised with I^2 [38]. Otherwise, synthesis followed SWiM without significance-based vote counting. Because no included report provided group-level means and standard deviations, an exploratory post hoc sensitivity analysis digitized the five available box-plot panels from the three SI-direct stool cohorts [9, 10, 13], converting digitized medians and interquartile ranges to approximate means and standard deviations (primary: mean = median, SD = IQR/1.349; sensitivity: mean = (Q1 + median + Q3)/3) [39, 40]. This analysis was sensitivity-only and not used for primary inference.

Author-provided data and post hoc validation. After the analysis cutoff, the authors of one SI-direct cohort [10] provided de-identified sample-level Shannon values for 141 participants (52 healthy controls, 32 depressed controls without SI, 57 with SI) and permission to use them in this report. These data were used only for two post hoc analyses: validation of the digitized medians and interquartile ranges, and a mixed exact/digitized sensitivity analysis that replaced the digitized values for this cohort while retaining digitized values for the other two cohorts. The pre-specified all-digitized analysis remains the registered

sensitivity analysis; the mixed analysis is exploratory and does not alter the primary synthesis, which was non-quantitative.

Risk of bias and certainty. Risk of bias was assessed with the JBI design-specific tools [41, 42, 43] and certainty with GRADE, starting at low for observational evidence [44]. Overall judgements (low, some concerns, high or critical) were derived from the domain-level JBI judgements using a pre-specified rule; "some concerns" denotes a JBI-based overall category, not the RoB 2 tool.

Part B: two-sample Mendelian randomization

The analysis followed STROBE-MR [45]; see Additional file 3.

Data sources. Instruments were derived from the MiBioGen GWAS of 18,340 participants from 24 cohorts with 16S data [26]; 131 of its 211 taxa were genera and were screened. The primary outcome was the European-ancestry suicide-attempt GWAS meta-analysis excluding the Utah cluster (EUR noutah; 31,094 cases, 758,690 controls) [4]. No cohort name was shared between the exposure and outcome samples, but PGC working-group entries are umbrella consortia with unlisted constituent cohorts, so individual-level overlap could not be excluded; overlap remains a potential bias [46]. Exposure-side replication used two Nordic metagenomic datasets: Dekkers (16,017 participants) and HUNT (12,652 participants) [47, 48].

Instruments and harmonisation. SNPs were selected at $P \leq 1 \times 10^{-5}$ and clumped ($r^2 < 0.001$ within 10 Mb; PLINK 1.9, 1000 Genomes Phase 3 European reference) [49]; SNPs with $F \leq 10$ were excluded [50]. Of 1,703 instrument rows, 1,619 (95.1%) were usable after harmonisation, with a median of 12 instruments per genus (range 4–24); 40 palindromic variants with minor allele frequency 0.42–0.58 and 44 variants absent from the outcome GWAS were removed.

Primary analysis and multiple testing. The MR estimate used inverse-variance weighted multiplicative random-effects (IVW-MRE) in TwoSampleMR 0.7.13 [51, 52]. Estimates are log odds ratios per 1 SD increase in genetically predicted genus abundance; ORs below 1 indicate a protective direction. The 131 genera formed a single fixed family with Benjamini–Hochberg control at $q < 0.05$ as the primary criterion and Bonferroni (3.82×10^{-4}) as the most conservative benchmark [53]; the same P values were also processed with Holm, Benjamini–Yekutieli and Storey procedures [54, 55]. Three testing families were kept separate: the 131-genus discovery family, a 39-entry replication family and a five-entry Gate R1 family.

Sensitivity, directionality, power and replication. For the FDR-significant genus, a pre-specified panel included weighted median MR, MR-Egger, MR-PRESSO, Cochran's Q and leave-one-out refitting [56, 57, 58]. Instrument pleiotropy was annotated across eight pre-specified domains (GWAS Catalog; $P \leq 5 \times 10^{-8}$ defining strong pleiotropy), and a pre-specified rule (Gate 4b) classified the signal as stable (A), suggestive (B) or not stable (C). Directionality used the Steiger test and reverse two-sample MR [59]. Minimum detectable effects (MDEs) were computed at 80% power and Bonferroni α [60, 61], interpreted

with the caveat that same-data instrument selection inflates apparent precision (winner's curse [62]). Exposure-side replication used a taxon-name crosswalk identifying 38 of 131 genera in the Nordic data (37 analysable; 39 entries). In replication step 1a, instrument–exposure associations were compared for direction concordance; in step 1b, instruments were selected within each Nordic dataset and analysed against the same suicide-attempt outcome (shared with discovery, so descriptive and not independent). A pre-specified Gate R1 classified entries as replicated (A), inconclusive (B) or incompatible (C).

Software. Analyses used R 4.6.1 (TwoSampleMR 0.7.13, PLINK 1.9, qvalue 2.44.0; renv, 181 packages). The primary plan was frozen internally before execution and the replication plan internally pre-registered; the study protocol was not publicly pre-registered.

Integration

The two parts were integrated descriptively: genera reported in Part A were matched by name to Part B estimates and q values, and classified as both streams with comparable direction, both streams without comparable direction, observational only, or genetic only. For this classification, a replicated observational association required the same taxon to be reported in the same direction in at least two independent cohorts; no taxon met this criterion. No formal triangulation, effect-size conversion, pooled analysis or difference test across parts was performed.

Results

Part A: study selection and characteristics

The searches identified 1,221 database and register records; 696 duplicates were removed, leaving 525 records for screening. A further 104 Google Scholar records were screened separately, giving 1,325 source records in total; 45 reports were assessed in full (Fig. 1). Thirty-two reports were excluded, leaving 13 reports from 11 independent cohorts (Fig. 1; Additional file 1: Table S1). Analysed sample sizes ranged from 46 to 236. Phenotype layers were SI-direct in three reports [9, 10, 13], SI-item in two [12, 63], continuous SI in two [64, 65], SA in one report using rectal swabs [16], suicide death in one report using post-mortem content [11] and proxy/mixed in four [14, 15, 17, 66]. Nine reports used 16S sequencing and six shotgun metagenomics; two used both.

Part A: extractability and alpha diversity

The 13 reports yielded 26 pre-specified Shannon/Chao1 synthesis units; **none contained group-level means, standard deviations, standard errors or confidence intervals**. Of 28 non-alpha narrative units, only one contained a numeric effect with a variance. No study reported a statistically significant Shannon difference for the target contrast. In the digitized sensitivity analysis of the three SI-direct cohorts, individual Hedges' g values for Shannon diversity were − 0.41 (SE 0.27) for Li et al. [13], + 0.24 (SE 0.22; digitized) for He et al. [10] and − 0.23 (SE 0.23) for Chen and Wu [9], giving a pre-specified

pooled estimate of − 0.11 (95% CI − 0.96 to 0.73; p = 0.62; I^2 = 50.2%; τ^2 = 0.058; Fig. 2); the alternative mean approximation gave − 0.16 (95% CI − 0.98 to 0.67). Author-provided sample-level data for He et al. [10] gave an exact g of 0.08 (95% CI − 0.35 to 0.52; 57 with SI and 32 without SI). Digitized medians and quartiles matched the exact values to within 0.002 units, but the median-as-mean approximation overestimated the effect in this cohort (0.24 versus 0.08); the alternative mean approximation gave 0.10. In a post hoc mixed exact/digitized analysis, the pooled estimate was − 0.16 (95% CI − 0.77 to 0.46; p = 0.38; I^2 = 7.7%; τ^2 = 0.005; Additional file 1: Table S5). For Chao1, two reports [9, 13] (g = − 0.43 and − 0.11) were not pooled (k = 2). Taxon-level findings were heterogeneous and non-overlapping (Additional file 1: Table S1).

Part A: risk of bias and certainty

Risk of bias was high in 10 reports, some concerns in two and critical in one (Fig. 3). Phenotype ascertainment was high risk in 10 of 13 reports, confounding in six, microbiota measurement in five, temporality in five and selective reporting in three. Certainty was very low for every principal question.

Part B: analysis set and instrument characteristics

Of 211 MiBioGen taxa, 131 genera were analysed (Additional file 1: Table S2). After harmonisation, 1,619 of 1,703 instrument rows (95.1%) were usable; the median number of instruments per genus was 12 (range 4–24). The primary outcome comprised 31,094 cases and 758,690 controls.

Part B: the single FDR signal

IVW-MRE estimates were centred close to the null (median β = −0.0048; negative point estimates for 72 genera, positive for 59; Fig. 4). Nine genera reached nominal P < 0.05, but only *Flavonifractor* survived BH-FDR correction (OR 0.819, 95% CI 0.737–0.910; P = 2.00×10^{-4}; q = 0.026; 8 instruments), which also met the Bonferroni threshold. The next-ranked genus, *Senegalimassilia* (OR 0.863, 95% CI 0.784–0.951; P = 0.0029), did not pass FDR correction (q = 0.191). Four of five correction procedures identified the same genus; the dependence-agnostic Benjamini–Yekutieli procedure identified none (BY q = 0.143), and $\hat{\pi}_0$ = 1.00 made Storey's q identical to the BH values.

Part B: sensitivity, directionality, power and replication

For *Flavonifractor*, sensitivity analyses were directionally consistent: weighted median OR 0.798 (95% CI 0.696–0.915; P = 0.0012), MR-Egger slope OR 0.912 (95% CI 0.586–1.420), no MR-PRESSO outlier (global P = 0.366), low heterogeneity (Q = 8.459, df = 7, P = 0.294; I^2 = 17.25%) and leave-one-out ORs of 0.793–0.839. One instrument, rs6761463, was strongly associated with body mass index (P = 2×10^{-14}) and adult body size (P = 1×10^{-11}); after removal, the estimate remained directionally consistent and

nominally significant (OR 0.830, 95% CI 0.738–0.933; P = 0.0018) but no longer met Bonferroni. The signal was therefore classified as suggestive (Gate 4b = B) rather than stable.

Steiger's test was consistent with an exposure-to-outcome direction for all 131 genera. Reverse MR found no genus surviving BH correction (minimum q = 0.354); for *Flavonifractor* the reverse estimate was close to the null (β = −0.021; P = 0.821). Median MDE was OR 0.861 (IQR 0.839–0.886) at the lower bound and OR 1.161 (IQR 1.129–1.193) at the upper bound; 126 of 131 genera could detect OR 0.80 and 109 OR 1.20, but only 15 could detect OR 0.90 and 6 OR 1.10.

In exposure-side replication, 37 of 131 genera (28.2%) were analysable and 39 entries entered the replication analysis. Instrument–exposure direction concordance was 18 of 39 entries (46.2%; binomial P = 0.749). Five entries had a discovery P < 0.05 and entered the pre-specified Gate R1 classification (A = 0, B = 3, C = 2); the remaining entries were descriptive. The primary replication entry, *Flavonifractor* genus abundance in Dekkers, gave OR 1.095 (95% CI 0.971–1.235; P = 0.141; power ≈ 0.998), in the opposite direction to discovery, and was classified R1-C; the species-level proxies were 0.954 (95% CI 0.862–1.055; R1-B) in Dekkers and 1.044 (95% CI 0.970–1.124; R1-C) in HUNT. In the separate 39-entry replication family, the smallest BH q was 0.549. Across alternative outcome versions, direction agreed with the primary analysis for 121 of 131 genera (92.4%) in the 2023 European-ancestry release excluding UK Biobank, 111 (84.7%) in the all-ancestry release and 108 (82.4%) in the 2022 release [67]; no genus survived FDR in any alternative version.

Integration: concordance and evidence gaps

Of the taxa reported narratively in Part A, nine could be matched by name to Part B estimates (Additional file 1: Table S3). None showed a robust genetic association: all matched estimates had q ≥ 0.844, and for most the observational findings were narrative directions without comparable group estimates, so direction concordance could not be formally evaluated. The only FDR-significant genus, *Flavonifractor*, was not reported in any included observational study, so the genetic signal has no direct observational counterpart in this corpus. The nominal genetic signals (*Lactococcus* OR 1.061, P = 0.027; *Peptococcus* OR 1.068, P = 0.034; *Hungatella* OR 1.097, P = 0.039; all q = 0.641) had no comparable genus-level observational data. No genus had both a replicated observational association and a robust genetic signal (Fig. 5).

Discussion

In this two-part study, the observational arm found that the current literature cannot support a protocol-defined meta-analysis of microbiota and suicidal behaviour: none of 13 reports from 11 cohorts provided group-level Shannon or Chao1 statistics. The pre-specified digitized sensitivity analysis gave a pooled Shannon difference of − 0.11 (95% CI − 0.96 to 0.73), and a post hoc analysis incorporating exact sample-level data for one cohort gave − 0.16 (− 0.77 to 0.46). This sensitivity analysis also showed that digitized medians and quartiles were accurate, whereas the median-as-mean conversion could bias the effect estimate. The genetic arm found a single suggestive genus, *Flavonifractor* (OR 0.819, 95% CI

0.737–0.910; q = 0.026), which did not replicate at the exposure side (OR 1.095; R1-C). No genus combined a replicated observational association with a robust genetic signal. Two designs with different bias structures therefore converge on the same conclusion: there is currently no robust evidence that gut microbiota is causally or reproducibly associated with suicidal behaviour, and the field's main deficit is evidence quality rather than a single missing study.

The observational findings extend previous syntheses: the updated corpus is larger and restricted to direct bacterial sequencing, however the conclusion remains suggestive but inconclusive, now grounded in a unit-by-unit account of what could and could not be estimated [18, 68]. Cross-diagnostic meta-analyses show that microbial differences can be detected when phenotypes are broad and samples large [7, 8]; the suicidal-behaviour literature has neither the sample size nor the phenotype standardisation to do the same.

The genetic findings agree with the broader microbiome-MR literature: instruments replicate poorly [28], power for small effects is limited [29], and previous studies of psychiatric outcomes yielded few signals [30, 31, 32, 33]. The *Flavonifractor* signal is best understood as a candidate that survived a large multiple-testing screen, was internally consistent across sensitivity and directionality analyses, but failed a pre-specified exposure-side replication and did not survive the dependence-agnostic correction. The replication data differ in measurement (shotgun versus 16S) and share the outcome with discovery, so the negative replication neither confirms nor refutes the discovery estimate.

The integration is deliberately modest. The two arms cannot be compared as two estimates of one causal parameter: Part A estimates associations between measured microbial composition and suicidal ideation, attempt or death in mostly Asian case–control samples; Part B estimates the effect of genetically predicted genus abundance on European-ancestry suicide attempt. Agreement would not prove causality and disagreement would not refute it; confounding, reverse causation, instrument invalidity, winner's curse, sample overlap and power each provide alternative explanations. What the integration provides is a map of where evidence exists: most narrative observational taxa have no robust genetic counterpart, the one robust genetic candidate has no observational counterpart, and several nominal genetic signals have not been examined observationally. This is a statement about the evidence base, not about biology.

Implications follow. Observational studies should report group-level means and standard deviations for every alpha-diversity index, deposit usable data accessions, pre-specify phenotype instruments and windows, and use phenotype-aligned psychiatric controls with reported residual symptoms. MR studies should pre-specify replication that changes the exposure measurement or the outcome sample, report instrument transfer concordance and power, and avoid describing single FDR signals as discoveries. For clinical readers, the combined evidence does not support microbiota-based tests, probiotics or dietary interventions for suicide risk assessment or prevention.

Limitations

Part A was restricted to English-language reports, included 13 reports and 11 cohorts, could not resolve cohort overlap, and had author-supplied data for only one of the 13 reports; the digitized sensitivity analysis is approximate and restricted to three cohorts, and the median-as-mean approximation overestimated the effect in the cohort with exact data, a bias that cannot be excluded for the other cohorts; publication bias could not be assessed; backward and forward citation searching and independent search-strategy review were not completed. Part B used instruments selected at $P \leq 1 \times 10^{-5}$ from 16S data, with limited variance explained and possible winner's curse; cohort-level overlap between exposure and outcome samples could not be excluded; only 38 of 131 genera were represented in the Nordic replication data (37 analysable); the outcome was shared between discovery and replication; power was limited for small-to-moderate effects; the analysis was restricted to European-ancestry data and sex- and gender-specific analyses were not possible because most reports did not stratify results by sex or gender; genus-level 16S relative abundance is compositional; and the instrumental-variable assumptions cannot be tested directly. Finally, the two parts estimate different quantities, so their combination is narrative and cannot resolve causal questions.

Conclusions

In this two-part study, the observational and genetic evidence streams converged on the absence of robust evidence for a microbiota–suicidal behaviour association. Two designs with different bias structures reaching the same boundary is informative about the state of the evidence, but it is not a causal finding and does not support clinical application. Phenotype-aligned, fully reported and adequately powered studies with pre-specified replication are needed before the field can move beyond this boundary.

Abbreviations

BH
Benjamini–Hochberg
BMI
body mass index
CC
case–control
CI
confidence interval
CO
cohort
CS
cross-sectional
FDR
false discovery rate
GWAS

genome-wide association study
GRADE
Grading of Recommendations, Assessment, Development and Evaluations
HUNT
Nord-Trøndelag Health Study
ISGC
International Suicide Genetics Consortium
IVW-MRE
inverse-variance weighted multiplicative random-effects
JBI
Joanna Briggs Institute
LOO
leave-one-out
MDD
major depressive disorder
MDE
minimum detectable effect
MR
Mendelian randomization
MR-PRESSO
Mendelian randomization pleiotropy residual sum and outlier
MVP
Million Veteran Program
NS
no suicidal ideation in major depressive disorder (comparator in Chen and Wu [9])
NSI
depressed controls without suicidal ideation (comparator in He et al. [10])
OR
odds ratio
PGC
Psychiatric Genomics Consortium
PRISMA
Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S
PRISMA search extension
q
Benjamini–Hochberg adjusted P value
RoB
risk of bias
SA

suicide attempt
SAP
statistical analysis plan
SC0
no current suicidal ideation
SC1
current suicidal ideation
SD
standard deviation
SI
suicidal ideation
SNP
single-nucleotide polymorphism
STROBE-MR
Strengthening the Reporting of Observational Studies in Epidemiology–Mendelian Randomization
SWiM
Synthesis Without Meta-analysis.

Declarations

Ethics approval and consent to participate

This study used published aggregate data (Part A), de-identified sample-level Shannon values provided by the authors of one included report, and summary-level genome-wide association data (Part B). No new participants were recruited and no new individual-level data were generated. Ethics approvals and consent procedures for the contributing studies are described in the original publications.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Guangzhou Research-oriented Hospital Program and by Grant No. 2024A03J0294 from the Guangzhou Municipal Science and Technology Project. The funders had no role in the design of the study, the collection, analysis or interpretation of data, or the writing of the manuscript.

Author Contribution

S.L. and X.L. conceived and designed the study. S.L. developed the search strategy, performed the searches, screened records, extracted data, assessed study quality, conducted the digitization, meta-analysis and Mendelian randomization analyses, prepared the figures and tables, and wrote the first draft. J.Z. independently screened records, extracted data, assessed study quality, and validated the analyses. Z.L. provided supervision and critically revised the manuscript. X.L. provided supervision, acted as the third reviewer for disagreements, and critically revised the manuscript. All authors read and approved the submitted version and agree to be accountable for all aspects of the work. X.L. is the guarantor.Abbreviations: S.L., Songqing Liu; J.Z., Jin Zhao; Z.L., Zezhi Li; X.L., Xiaobing Lu.

Acknowledgement

The authors thank the authors of He et al. [10] for providing sample-level Shannon data with permission; the MiBioGen consortium; the Psychiatric Genomics Consortium and the International Suicide Genetics Consortium; the Million Veteran Program; the Dekkers and HUNT study teams; the GWAS Catalog; and the contributing observational study teams.

Data Availability

No new individual-level data were generated. Aggregated data supporting the findings are included in the article and Additional file 1 (Tables S1–S5). The full extraction tables, availability register, digitization calibration record and analysis scripts are available from the corresponding author on reasonable request.Part A: all 13 included reports are published. De-identified sample-level Shannon data for He et al. [10] were provided by the original authors with permission; they are not publicly deposited, and requests should be directed to the original authors.Part B: MiBioGen summary statistics are publicly available from the MiBioGen consortium (https://mibiogen.gcc.rug.nl/) [26]. Suicide-attempt GWAS summary statistics from Docherty et al. [4] are available through the PGC Suicide Working Group Data Access Portal (https://pgc.unc.edu/for-researchers/working-groups/suicide-working-group/) under a data use agreement and cannot be redistributed. Nordic replication summary statistics from Dekkers et al. [47] (https://doi.org/10.1038/s41588-026-02512-2) and Moksnes et al. [48] (https://doi.org/10.1038/s41588-026-02502-4) were obtained from the GWAS Catalog (https://www.ebi.ac.uk/gwas/) under CC0 terms. No new individual-level data were generated.

References

1. Fazel S, Runeson B, Suicide. N Engl J Med. 2020;382(3):266–74.
2. Turecki G, Brent DA. Suicide and suicidal behaviour. Lancet. 2016;387(10024):1227–39.

3. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull. 2017;143(2):187–232.
4. Docherty AR, Mullins N, Ashley-Koch AE, Qin X, Coleman JRI, Shabalin A, et al. GWAS meta-analysis of suicide attempt: identification of 12 genome-wide significant loci and implication of genetic risks for specific health factors. Am J Psychiatry. 2023;180(10):723–38.
5. Cryan JF, O'Riordan KJ, Cowan CSM, Sandhu KV, Bastiaanssen TFS, Boehme M, et al. The microbiota-gut-brain axis. Physiol Rev. 2019;99(4):1877–2013.
6. Sharon G, Sampson TR, Geschwind DH, Mazmanian SK. The central nervous system and the gut microbiome. Cell. 2016;167(4):915–32.
7. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol. 2019;4(4):623–32.
8. Nikolova VL, Smith MRB, Hall LJ, Cleare AJ, Stone JM, Young AH. Perturbations in gut microbiota composition in psychiatric disorders: a review and meta-analysis. JAMA Psychiatry. 2021;78(12):1343–54.
9. Chen VC, Wu SI. An exploratory analysis on the association between suicidal ideation and the microbiome in patients with or without major depressive disorder. J Affect Disord. 2025;370:362–72.
10. He L, Huang Y, Li H, Zhu B, Zhang Z, Wu J, et al. Novel insights into gut microbiota alterations in major depressive disorder with suicidal ideation: a metagenomic analysis. Front Microbiol. 2026;17:1843301.
11. Kozhakhmetov S, Kossumov A, Zhakupova T, Polyakova T, Imambayeva N, Syzdykova B, et al. Characterization of gut microbiome composition in depression and completed suicide. Int J Mol Sci. 2025;26(10):4880.
12. Liang L, Chen S, Zhu B, Huang Y, Lu X, Wang S, et al. Continuous alterations in the gut microbial landscape associated with suicidal ideation in first-episode drug-naïve major depressive disorder. CNS Neurosci Ther. 2026;32(6):e70892.
13. Li F, Jiang Z, Xing J, Guo F, Qu C, Zhang A, et al. Characterization and correlation of gut microbiota in major depressive disorder patients with suicidal ideation. BMC Psychiatry. 2026;26(1):738.
14. Maes M, Vasupanrajit A, Jirakran K, Klomkliew P, Chanchaem P, Tunvirachaisakul C, et al. Adverse childhood experiences and reoccurrence of illness impact the gut microbiome, which affects suicidal behaviours and the phenome of major depression: towards enterotypic phenotypes. Acta Neuropsychiatr. 2023;35(6):328–45.
15. Maes M, Zhou B, Vasupanrajit A, Jirakran K, Klomkliew P, Chanchaem P, et al. A further examination of growth factors, T helper 1 polarization, and the gut microbiome in major depression: associations with reoccurrence of illness, cognitive functions, suicidal behaviors, and quality of life. J Psychiatr Res. 2024;176:430–41.

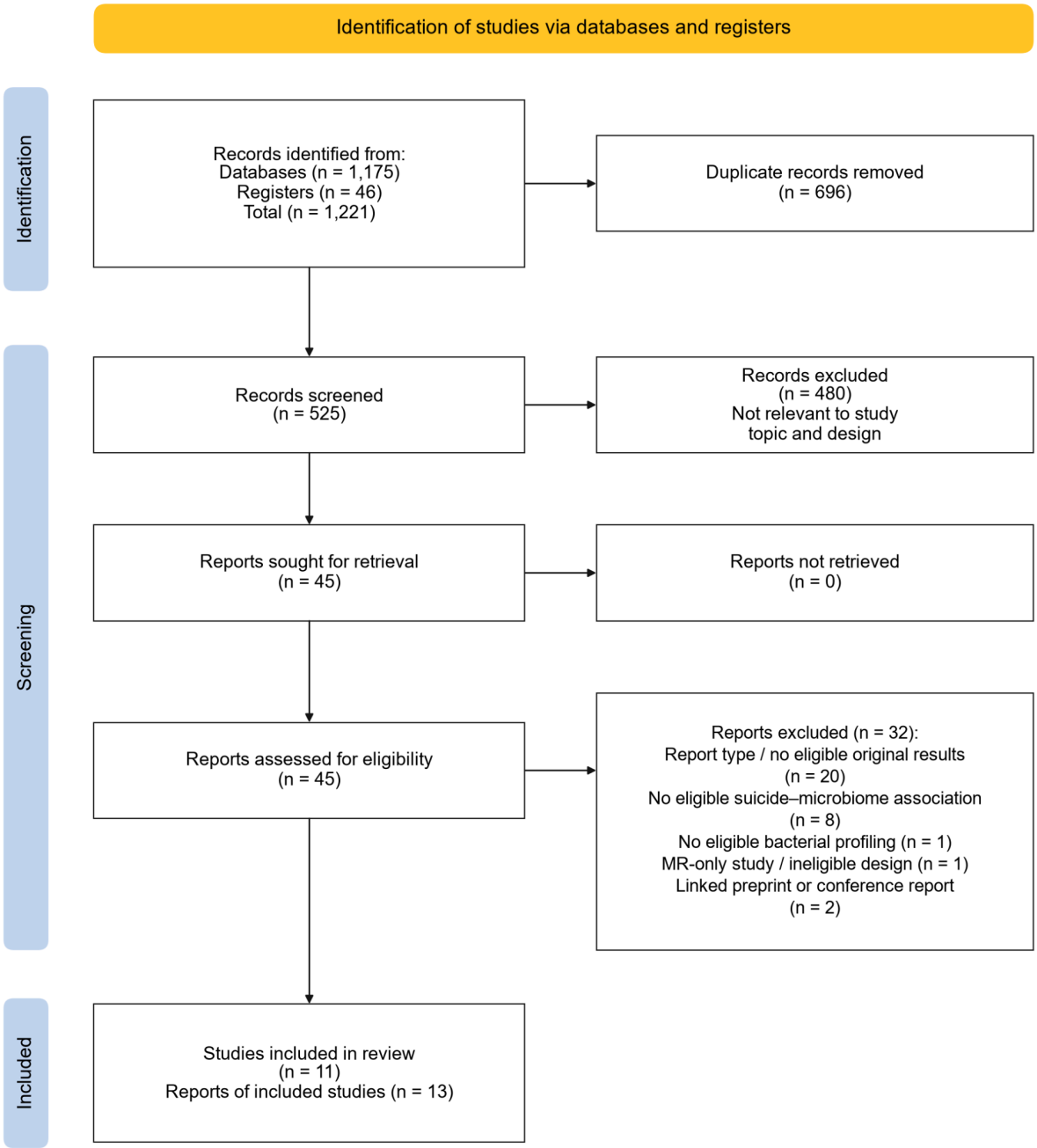
16. Prandovszky E, Liu H, Severance EG, Splan VW, Dickerson FB, Yolken RH. Altered gut microbial diversity, composition, and metabolomic potential in patients with major depressive disorder and recent suicide attempt. Brain Behav Immun Health. 2025;48:101081.
17. Thompson DS, Fowler JC, Bradshaw MR, Frueh BC, Weinstein BL, Petrosino J, et al. Is the gut microbiota associated with suicidality? Non-significant finding among a large cohort of psychiatrically hospitalized individuals with serious mental illness. J Affect Disord Rep. 2021;6:100266.
18. Zerekidze A, Li M, Refisch A, Shameya J, Sobanski T, Walter M, et al. Impact of Toxoplasma gondii and human microbiome on suicidal behavior: a systematic review. J Clin Med. 2024;13(2):593.
19. Baldini V, Gnazzo M, Santangelo G, Varallo G, De Ronchi D, Carotenuto M. The role of gut microbiota in suicidality: mechanisms, evidence, and future directions. Psychiatry Int. 2025;6(3):84.
20. Shamabadi A, Arabzadeh Bahri R, Arab Bafrani M, Karimi H, Asadigandomani H, Vahidi H, et al. The gut-brain axis and inflammatory mediators in suicide and mental disorders with high suicide rates: a review of current evidence. Dialogues Clin Neurosci. 2026;28(1):145–56.
21. Na PJ, Yaramala SR, Kim JA, Kim H, Goes FS, Zandi PP, et al. The PHQ-9 item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)-9 item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). J Affect Disord. 2018;232:34–40.
22. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. Sci Rep. 2021;11(1):3030.
23. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
24. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601.
25. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, et al. Mendelian randomization. Nat Rev Methods Primers. 2022;2:6.
26. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet. 2021;53(2):156–65.
27. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019;51(4):600–5.
28. Hughes DA, Bacigalupe R, Wang J, Rühlemann MC, Tito RY, Falony G, et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat Microbiol. 2020;5(9):1079–87.
29. Wade KH, Hall LJ. Improving causality in microbiome research: can human genetic epidemiology help? Wellcome Open Res. 2020;4:199.

30. Chen M, Xie CR, Shi YZ, Tang TC, Zheng H. Gut microbiota and major depressive disorder: a bidirectional Mendelian randomization. J Affect Disord. 2022;316:187–93.
31. Ni JJ, Xu Q, Yan SS, Han BX, Zhang H, Wei XT, et al. Gut microbiota and psychiatric disorders: a two-sample Mendelian randomization study. Front Microbiol. 2022;12:737197.
32. Gao J, Pan L, Wang X, Xu J, Xu W, Zhang Y, et al. Causal relationship between gut microbiota and non-suicidal self-injury: a two-sample Mendelian randomization study. Am J Med Genet B Neuropsychiatr Genet. 2025;198(8):241–50.
33. Yu T, Chen C, Yang Y, Wang M, Yang Y, Feng W, et al. Dissecting the association between gut microbiota, body mass index and specific depressive symptoms: a mediation Mendelian randomisation study. Gen Psychiatr. 2024;37(4):e101412.
34. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ. 2020;368:l6890.
35. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
36. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39.
37. Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. BMC Med Res Methodol. 2015;15:99.
38. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557–60.
39. Luo D, Wan X, Liu J, Tong T. Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. Stat Methods Med Res. 2018;27(6):1785–805.
40. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol. 2014;14:135.
41. Barker TH, Hasanoff S, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for cohort studies. JBI Evid Synth. 2025;23(3):441–53.
42. Barker TH, Hasanoff S, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for analytical cross-sectional studies. JBI Evid Synth. 2026;24(3):401–8.
43. JBI. JBI manual for evidence synthesis (Chap. 7: systematic reviews of etiology and risk). JBI. 2024. https://doi.org/10.46658/JBIMES-24-06. Accessed 4 November 2026.
44. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924–6.

45. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian randomization: the STROBE-MR statement. JAMA. 2021;326(16):1614–21.
46. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597–608.
47. Dekkers KF, Pertiwi K, Baldanzi G, Lundmark P, Hammar U, Moksnes MR, et al. Genome-wide association analyses highlight the role of the intestinal molecular environment in human gut microbiota variation. Nat Genet. 2026;58(3):540–9.
48. Moksnes MR, Coward E, Nethander M, Dekkers K, Grahnemo L, Törnqvist AE, et al. The HUNT study identifies host genetic factors reproducibly associated with human gut microbiota composition. Nat Genet. 2026;58(3):530–9.
49. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.
50. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol. 2011;40(3):755–64.
51. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658–65.
52. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
53. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol. 1995;57:289–300.
54. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001;29:1165–88.
55. Storey JD. A direct approach to false discovery rates. J R Stat Soc Ser B Stat Methodol. 2002;64:479–98.
56. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.
57. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol. 2016;40(4):304–14.
58. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693–8.
59. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13(11):e1007081.
60. Brion MJA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497–501.

61. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol. 2014;43(3):922–9.
62. Jiang T, Gill D, Butterworth AS, Burgess S. An empirical investigation into the impact of winner's curse on estimates from Mendelian randomization. Int J Epidemiol. 2023;52(4):1209–19.
63. Wang W, Zhu B, Huang Y, Liang L, Li Z, Diao Y, et al. Analysis of individualized brain–gut association networks and transcriptional patterns in major depressive disorder. Interdiscip Med. 2026. 10.1002/inmd.70165.
64. Maes M, Almulla AF, Vasupanrajit A, Jirakran K, Tunvirachaisakul C, Maes A, et al. Functional shotgun metagenomic insights into gut microbial pathway and enzyme disruptions linking metabolism, affect, cognition, and suicidal ideation in major depressive disorder. Acta Neuropsychiatr. 2026;38:e16.
65. Meng C, Li J, Niu M, Almulla AF, Luo Y, Carvalho AF, et al. Systems biology of gut microbiome dysbiosis in major depressive disorder: functional microbial networks linked to inflammation, illness severity, recurrence, and suicidal behavior. Res Square [Preprint]. 2026. 10.21203/rs.3.rs-10598930/v1.
66. Xu S, Xiong J, Qin X, Ma M, Peng Y, Cheng J, et al. Association between gut microbiota and perinatal depression and anxiety among a pregnancy cohort in Hunan, China. Brain Behav Immun. 2025;125:168–77.
67. Mullins N, Kang J, Campos AI, Coleman JRI, Edwards AC, Galfalvy H, et al. Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. Biol Psychiatry. 2022;91(3):313–27.
68. Costanza A, Alexander J, Amerio A, Aguglia A, Magnani L, Parise A, et al. Microbial dysbiosis as an emerging pathology of suicidal behaviour? A critical review, passing through depression to chronic pain. Ann Neurosci. 2026;33(3):365–79.

Figures



Google Scholar: 104 records screened separately; stage-specific counts were not retained.
Adapted from PRISMA 2020 (Page et al., BMJ 2021;372:n71).

Figure 1

PRISMA 2020 flow diagram. Databases and registers identified 1,221 records; after removal of 696 duplicates, 525 records were screened and 45 reports were assessed in full, of which 32 were excluded and 13 reports (11 independent cohorts) were included. A further 104 Google Scholar records were screened separately; stage-specific counts for that source were not retained.

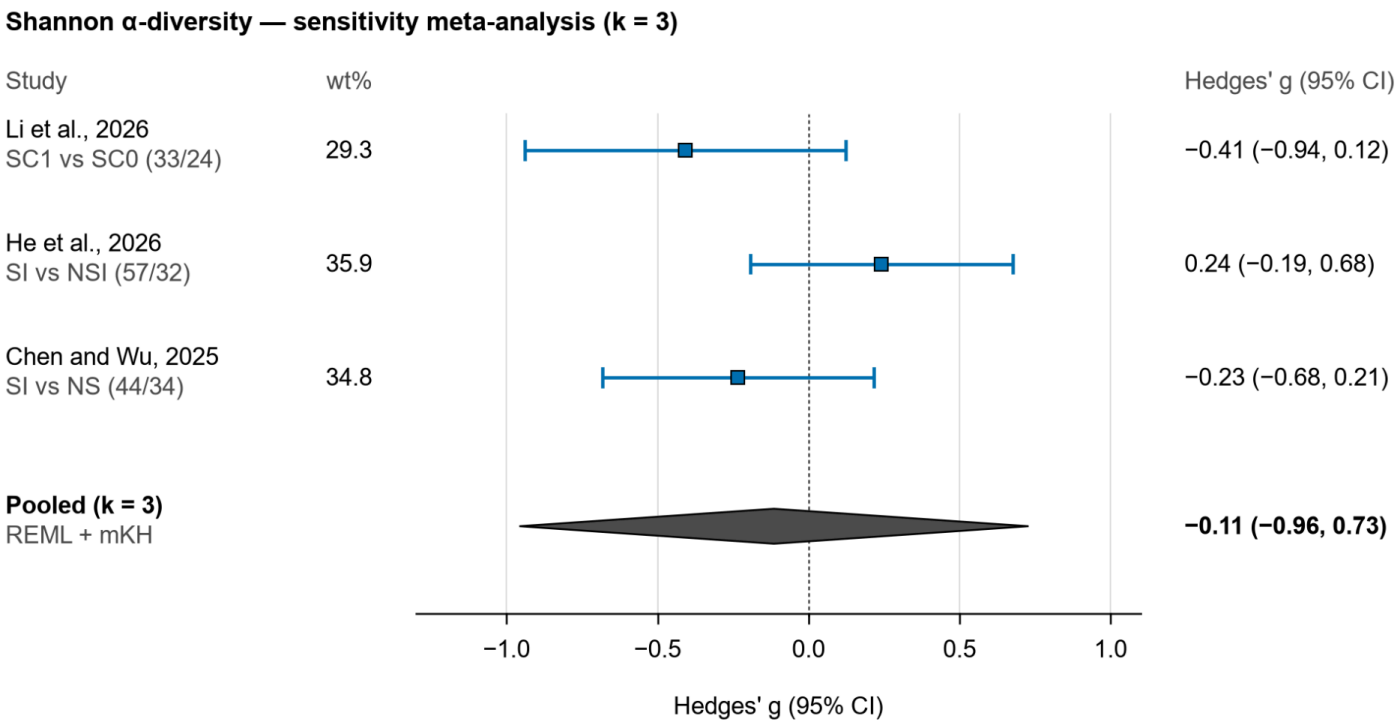


Figure 2

Digitized Shannon diversity sensitivity analysis. Exploratory graph-digitized analysis in three SI-direct cohorts (Li et al., 2026; He et al., 2026; Chen and Wu, 2025). SC1, current suicidal ideation; SC0, no current suicidal ideation; SI, suicidal ideation; NSI, no suicidal ideation; NS, no suicidal ideation in major depressive disorder. The pre-specified all-digitized estimate was g = −0.11 (95% CI −0.96 to 0.73; I^2 = 50.2%). In a post hoc analysis using author-provided sample-level data for He et al. (2026), the estimate was −0.16 (95% CI −0.77 to 0.46; I^2 = 7.7%). The analysis is sensitivity-only: means and standard deviations were approximated from digitized box plots for two cohorts, no primary meta-analysis was possible, and the estimate must not be interpreted as evidence of no difference. Validation values are reported in Additional file 1: Table S4 and the mixed analysis in Additional file 1: Table S5.

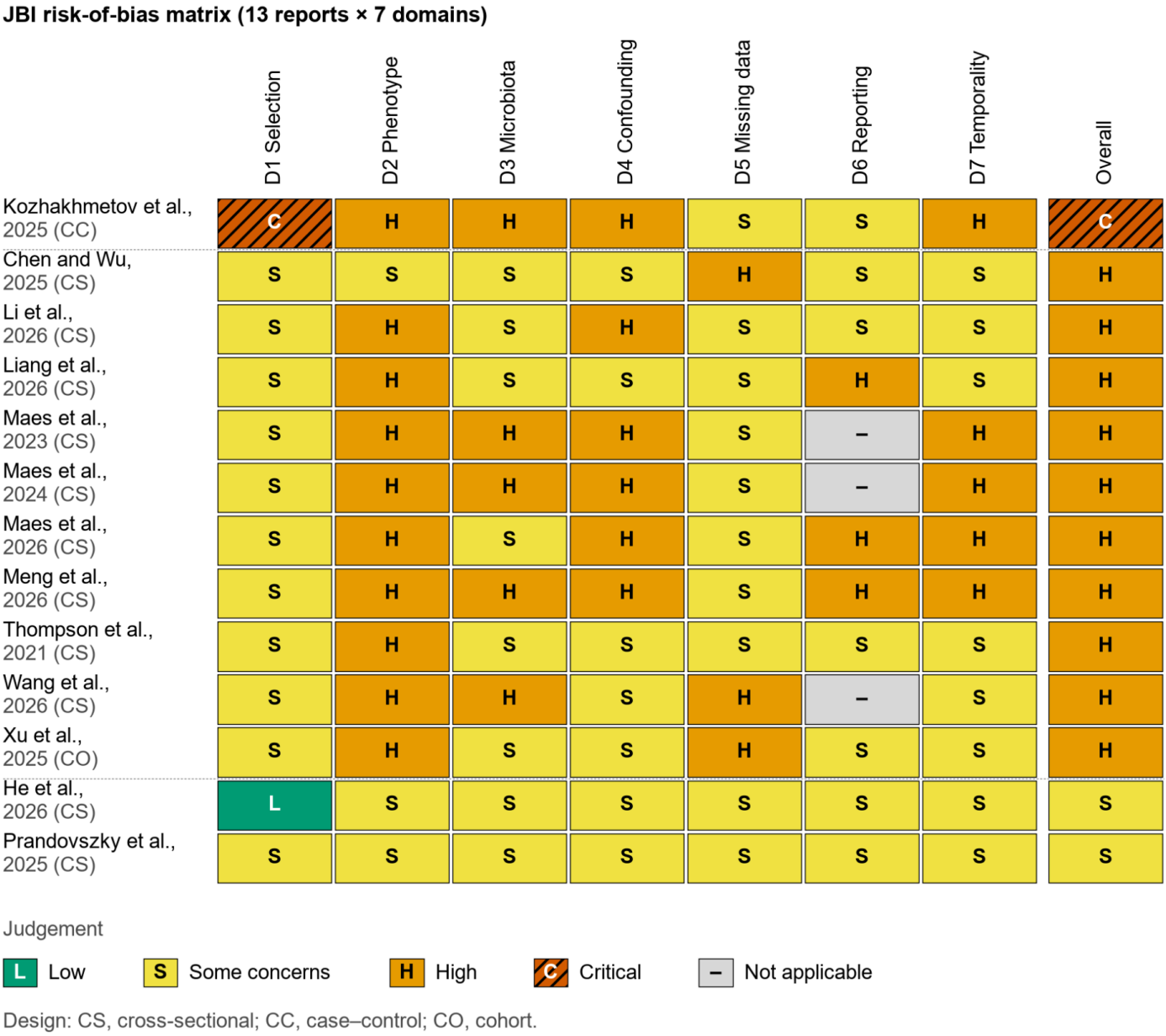


Figure 3

Risk of bias in the included observational reports. Domain-level judgements for the 13 reports in Part A using JBI design-specific domains, with overall judgements. Judgements were made independently by two reviewers; disagreements were resolved by consensus. CS, cross-sectional; CC, case–control; CO, cohort.

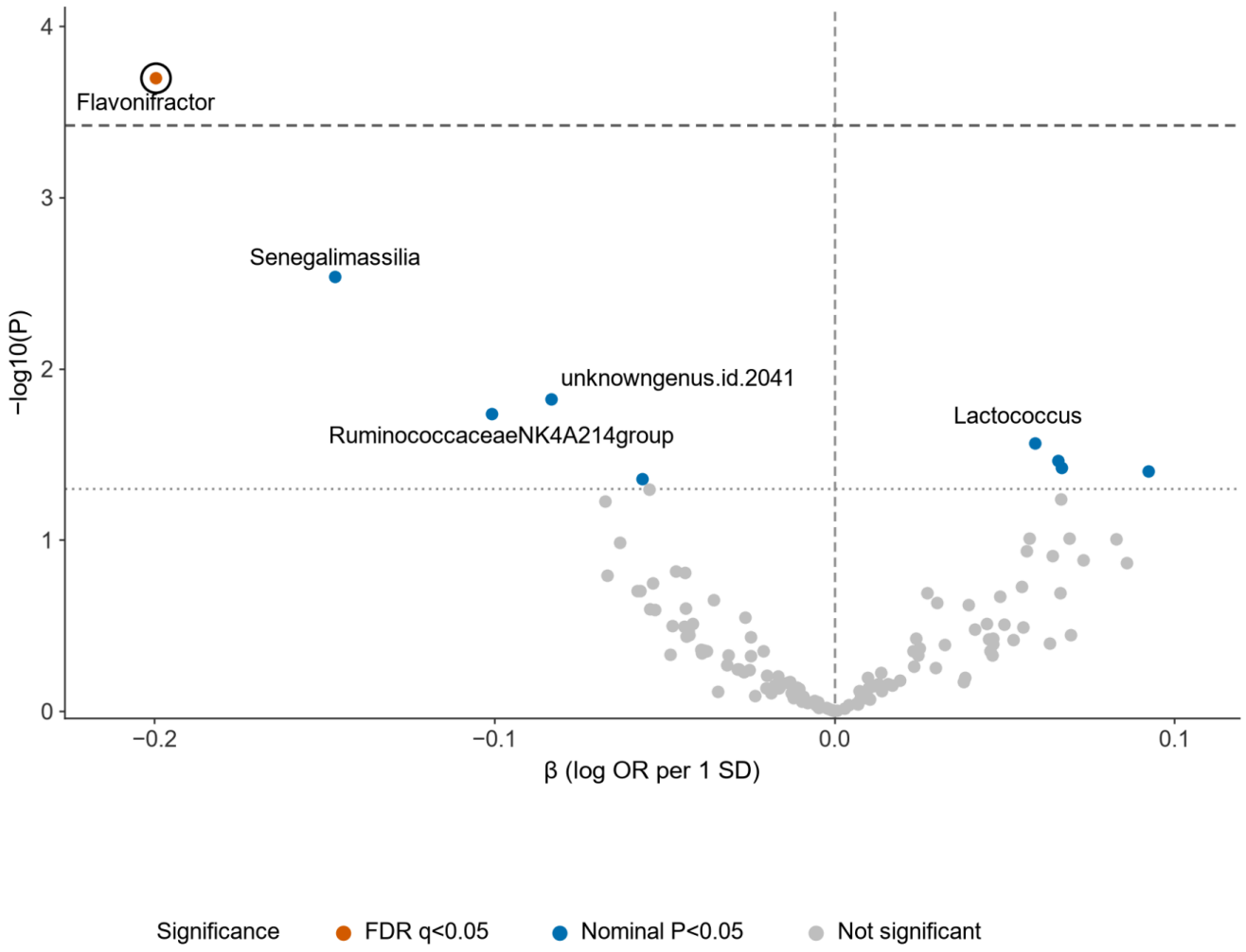


Figure 4

Mendelian randomization screen of 131 bacterial genera. Volcano plot of the 131-genus inverse-variance weighted multiplicative random-effects analysis of suicide attempt. The x-axis shows the point estimate (log OR) and the y-axis the −log10 P value; the dotted horizontal line marks nominal P = 0.05 and the dashed horizontal line marks the Bonferroni threshold ($P = 3.82 \times 10^{-4}$; 0.05/131). *Flavonifractor* is the only genus reaching FDR significance (OR 0.819, 95% CI 0.737–0.910; $P = 2.00 \times 10^{-4}$; q = 0.026); the other eight genera with P < 0.05 are shown but do not pass FDR correction. Estimates are ORs per 1 SD genetically predicted genus abundance, and ORs below 1 indicate a protective direction.

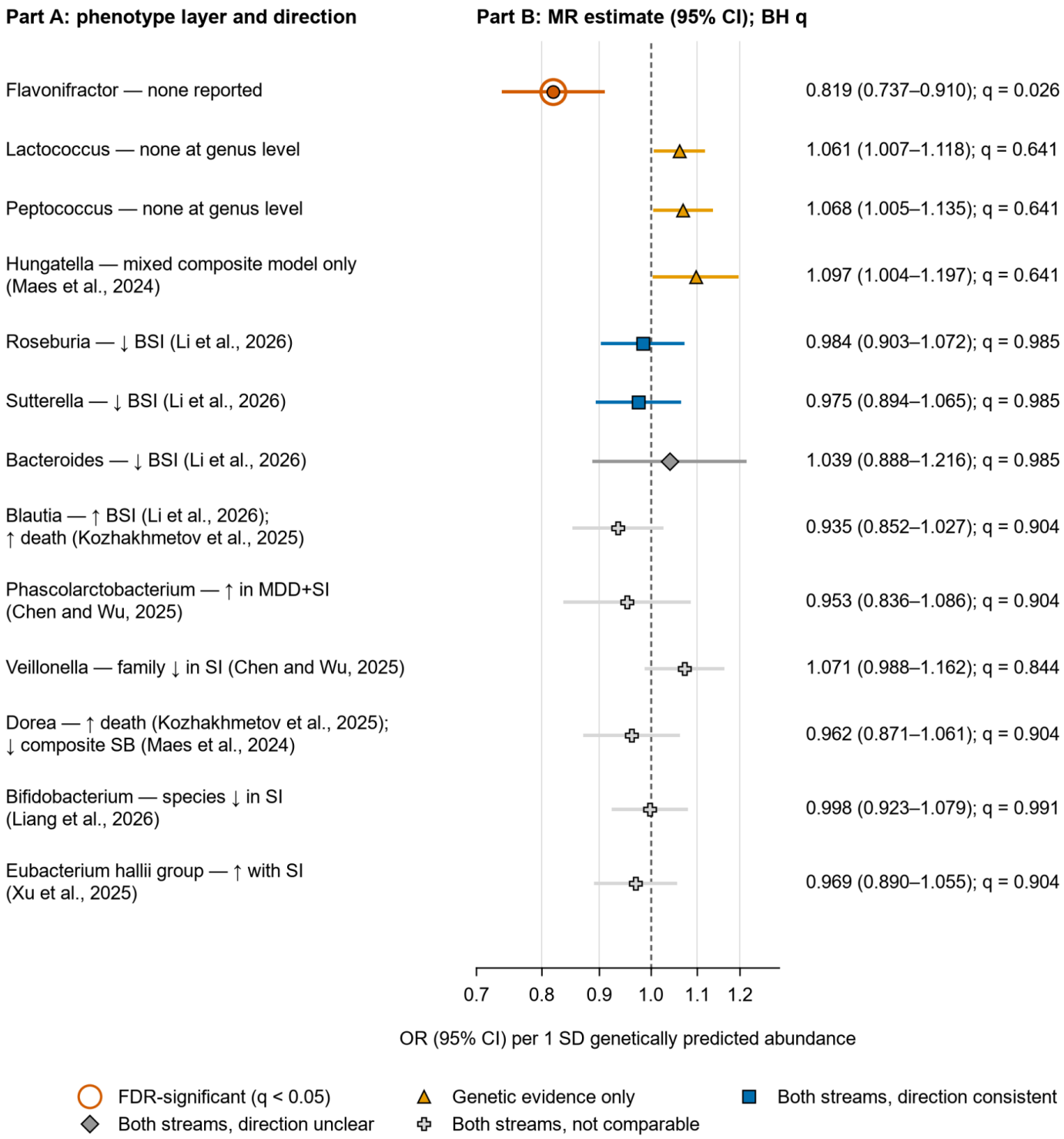


Figure 5

Descriptive integration of observational and genetic evidence. For each genus, the left column gives the Part A phenotype layer and reported direction, the middle panel shows the Part B Mendelian randomization OR with 95% confidence interval (log scale), and the right column gives the Benjamini–Hochberg q value. Symbols and colours distinguish FDR-significant, genetic-evidence-only, both streams with consistent direction, both streams with unclear direction and both streams not comparable.

*Flavonifractor*is the only FDR-significant genus and had no observational counterpart; no genus combined a replicated observational association with a robust genetic signal. The figure is descriptive and is not a formal triangulation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx