




METHODS PAPER

Risk of bias assessment tool for systematic review and meta-analysis of the gut microbiome

Thomas Lampeter¹ , Charles Love², Trien T. Tang², Aditi S. Marella², Hayden Y. Lee², Armani Oganyan², Devin Moffat², Anisha Kareem², Matthew Rusling² , Aubrey Massmann², Melanie Orr¹, Christian Bongiorno² and Li-Lian Yuan² 

¹New York Institute of Technology College of Osteopathic Medicine, Glen Head, NY, USA

²Des Moines University College of Osteopathic Medicine, Des Moines, IA, USA

Corresponding authors: Thomas Lampeter, Matthew Rusling, and Li-Lian Yuan; Emails: Lampeter.thomasm@gmail.com; matthewrusling@gmail.com; lilian.yuan@dmu.edu

(Received 30 September 2022; revised 10 April 2023; accepted 01 July 2023)

Abstract

Risk of bias assessment is a critical step of any meta-analysis or systematic review. Given the low sample count of many microbiome studies, especially observational or cohort studies involving human subjects, many microbiome studies have low power. This increases the importance of performing meta-analysis and systematic review for microbiome research in order to enhance the relevance and applicability of microbiome results. This work proposes a method based on the ROBINS-I tool to systematically consider sources of bias in microbiome research seeking to perform meta-analysis or systematic review for microbiome studies.

Keywords: bias; microbiome; metaanalysis; systematic review

Introduction

The most common experimental design used to evaluate the effects of gut microbiome (GMB) genomic or taxonomic post-exposure remodelling has been cohort studies using either animal or human models. Randomised controlled trials (RCTs) for microbiome interventions are less common because we are still characterising microbiome post-exposure remodelling to identify promising markers or targets for microbiome intervention that would warrant subsequent evaluation by RCTs. Therefore, results from a systematic review with quantitative or pooled meta-analysis are essential in identifying candidates for RCTs.

A diligent risk of bias (ROB) assessment is a key step in systematic review or meta-analysis to determine the likelihood that features of the study design or conduct of the study will give misleading results. GMB research is highly heterogeneous in its methods, reporting, and attempts to address bias. This manuscript and its associated rubric (Table 1) are based on the Risk of Bias in Non-randomised Studies – of Interventions (ROBINS-I) tool, and are meant to be used as a GMB-specific adjunct to ROBINS-I. This manuscript and its associated rubric together form a tool that was developed to help standardise ROB assessment in meta-analyses and systematic reviews of GMB studies. A small-scale validation test by first-time ROB assessors produced consistently similar ROB determinations, suggesting that this tool can successfully guide consistent ROB determinations. This tool may allow for improved ROB assessment when evaluating studies for meta-analyses and systematic reviews of the GMB.

Table 1. The rubric of domains and subdomains of bias with signalling statements to guide risk of bias assessment of gut microbiome studies.

Domain	High ROB	Moderate ROB	Low ROB
Confounding			
1. Demographic differences	<ul style="list-style-type: none"> – Age: consistently different between study arms – Sex: consistently different between study arms 	<ul style="list-style-type: none"> – Age: mixed ages within a study arm, but equal in distribution between study arms – Sex: mixed sexes within a study arm, but equal in distribution between study arms 	<ul style="list-style-type: none"> – Age: consistently similar between study arms – Sex: consistently similar between study arms
2. Habitat stability	<ul style="list-style-type: none"> – No acclimation period, or acclimation period <2 days – Acclimation period included in the interventional period 	<ul style="list-style-type: none"> – Acclimation period ≥ 2 days but <5 days 	<ul style="list-style-type: none"> – Acclimation period ≥ 5 days and < 9 weeks
3. Genotype, familial, and source differences	<ul style="list-style-type: none"> – Significantly different subject genotypes between study arms (where genotype effect is not the target of investigation) – Non-matured animal models from different litters and/or mothers without random assortment into study arms – Comparison of animal subjects from different source or vendor between study arms 	<ul style="list-style-type: none"> – Animal subjects from same vendor, but from separate and temporally spaced orders without random assortment into study arms 	<ul style="list-style-type: none"> – Adequately similar genotypes used between study arms (where host genotype effect is not the target of study) – Animal subjects from same litter – Animal subjects from same vendor and same order – Adult animal subjects from different litters/ mothers/vendors randomly assorted into study arms
4. Extreme diet	<ul style="list-style-type: none"> – No statement of dietary standards or documentation of dietary variation – Major deviations from stated diet 	<ul style="list-style-type: none"> – Study uses human subjects outside of a highly controlled environment (for example an inpatient healthcare setting) 	<ul style="list-style-type: none"> – Use of identical diet between study arms where diet is not the target of study
5. GMB normalisation	<ul style="list-style-type: none"> – No documented means of verified GMB normalisation methods employed prior to intervention – Use of different normalisation methods between study arms or use of non-validated technique 	<ul style="list-style-type: none"> – Antibiotic normalisation employed <7 days prior to intervention 	<ul style="list-style-type: none"> – Antibiotic normalisation employed ≥ 7 days prior to intervention – Validated technique of GMB normalisation employed
Selection Bias			
1. Extreme genotype	<ul style="list-style-type: none"> – Subjects of known extremely different genotypes – Subjects with no established history of use 	<ul style="list-style-type: none"> – Syngeneic subjects with limited established history of use 	<ul style="list-style-type: none"> – Syngeneic subjects with established history of use

Table 1. *Continued*

Domain	High ROB	Moderate ROB	Low ROB
2. Randomisation or demographic balancing sufficiently applied	– Absence of both RCT and implementation of consistent host demographic across study	– Utilisation of RCT or implementation of consistent host demographics across study	– Utilisation of RCT and implementation of consistent host demographics across study
Classification of Intervention			
1. Intervention bias	– Differential misclassification of intervention or test subject based on exposures present or suspected	– n/a	– Differential misclassification of intervention or test subject based on exposures absent or not suspected
2. Validation of method	– No validation that treatment method produces intended effect – Use of new method without internal validation	– n/a	– Documented use of validated methods – Use of a new method with adequate internal validation
Deviation from Intervention			
1. Deviation from intervention	– Large deviations to protocol without adequate time stamps, rationale, and limitations noted	– Slight deviations to protocol with adequate time stamps, rationale, and limitations noted	– Intervention successfully carried out without protocol deviation
Missing Data			
1. Cause or category of missing data	– Does not address missing data qualitatively or quantitatively – Or, does not ensure to readers the absence of missing data	– Acknowledges missing data qualitatively or quantitatively – Inadequate MAR/MNAR distinction or proper statistical correction	– Addresses missing data qualitatively or quantitatively, or ensures the absence of missing data – Adequate MAR/MNAR distinction or proper statistical correction
2. Subject dropout	– Subject drop-out exceeds 20%	n/a	– Subject drop-out is equal to or less than 20%
3. Sequencing depth and sampling zeroes	– Does not address sampling zeroes with statistical correction	n/a	– Addresses sampling zeroes with statistical correction
Measurement of Outcomes			
1. Sample collection	– Inconsistent collection time	– Animal models: Collected at same time, not in the morning – Human models: Inconsistent collection time, but preserved immediately after defecation	– Animal models: Collected at same time, in the morning – Human models: Consistent collection time and preserved immediately after defecation

Table 1. *Continued*

Domain	High ROB	Moderate ROB	Low ROB
2. Blinding	– No double blinding when the primary measurement is subjective	n/a	– Primary outcome is objective measure such as genetic sequencing not subject to bias by the subject or investigator – Primary outcome is subjective and double or greater blinding employed
Reporting of Results			
1. Selection of reported results	Any of: – Omission of stated outcomes that are unfavourable or statistically insignificant – Addition of outcomes not in the initial protocol – Results reported are only on a subset of data – Changing outcome(s) of interest	– Any of the above, but with valid and satisfactory explanation provided	– Inclusion of relevant null and significant findings as stated in the protocol

Using this tool

This manuscript and its associated rubric provide a framework for assessing ROB specific to GMB research. This tool strives to provide insight and reduce variability between individual researchers and groups conducting systematic reviews of the GMB. We do not seek to suggest best practices. Instead, we aim to indicate potential sources of bias that may significantly impact GMB studies and are thus vital when considering the strength of evidence for systematic review and meta-analysis. The essential criteria in this manuscript are summarised in [Table 1](#), which was compiled to act as a rubric in guiding ROB determination.

[Table 1](#), “the rubric,” guides the determination of low, moderate, or high ROB across seven domains. In each cell of the rubric, there are signalling statements to help guide low, moderate, or high ROB determination in that domain. Two additional ROB determinations are not included on the rubric as they are to be used at the judgement of the person assessing ROB in a study. They are “critical ROB” and “no information.” Critical ROB can be determined when a reviewer believes a study to be too problematic to provide useful evidence on the effect of an intervention. As such, a study determined to be of critical ROB in any one domain should not be included in any synthesis. A determination of no information applies to domains where there is no clear evidence of a critical ROB *and* a lack of information to judge ROB otherwise.

Confounding

Demographic differences

Important demographic considerations in GMB studies are sex and age. Substantial differences in the gut microbiota are attributable to sex differences in mammals (Org et al., 2016; Kim et al., 2020). Because of this, any study which includes one sex in one arm and a different sex in another should be classified as having a high ROB. In addition to the ROB from sex, other demographic factors may also introduce confounding bias into the studies being examined. The GMB changes with age across numerous conditions, disease models, and species impacting microbial diversity and biome composition (Ticinesi et al., 2019; Liu et al., 2020). Therefore, age differences between cohorts and study arms should be assessed. If the study being examined uses organisms of one age in one arm and a different age in a second arm, it should be classified as having a high ROB. The age gap which introduces significant confounding bias, varies by organism. An example of an age gap that would introduce a high ROB is 8-week-old mice versus 1-year-old mice (Yoon et al., 2021).

Habitat stability

The habitat in which organisms are kept substantially impacts their GMB (Singh et al., 2021). Mice, common subjects of microbiome research, are known to have highly variable microbiomes on arrival at a facility, likely because of transportation stress on the microbiome itself and the immune system and hormonal functions of the host organism (Capdevila et al., 2007; Montonye et al., 2018; Lipinski et al., 2021). Studies that do not allow for microbiome stabilisation before research begins risk confounding bias due to a lack of habitat stability. Organisms should be acclimated to the study condition before baseline measurements or interventions are performed. However, an extensive acclimation period risks microbiome drift occurring due to the increasing age of the organism or other unknown factors, so habitat stabilisation must be time-limited (Hoy et al., 2015). Additional bias would also be introduced if the acclimation period is included in the interventional period of the research.

Genotype, familial, and source differences

Subject genotype, degree of familial relation, and in the case of animal models, the source can significantly impact GMB composition. Differences in the genotype of animal models have been found

to impact the diversity and abundance of organisms (Campbell et al., 2012; McKnite et al., 2012; Leamy et al., 2014). For this reason, if the study being evaluated uses organisms of significantly different genotypes, such as the use of different strains of mice from the Collaborative Cross, where the effect of genotype difference is not the target of the study, it should be classified as having a high ROB. Suppose the study uses a similar genotype between treatment groups, such as the same strain of inbred animal model or monozygotic twin subjects. In that case, it should be considered a low ROB for confounding due to the genotype effect.

Regarding familial relation, genetically related subjects have been demonstrated to share a core of similar GMB for up to three generations in the female line (Turnbaugh et al., 2008; Valles-Colomer et al., 2021). With animal models, breeding within familial relations is often used to maintain genotypically and GMB homogeneity (Hufeldt et al., 2010). A caution regarding inbreeding is that while selective breeding between siblings can create a more stable and uniform GMB composition, the effects of genetic drift can also introduce confounders across multiple generations that may affect experimental reproducibility with subsequent generations (Laukens et al., 2016).

Additionally, with animal models, an organism's litter of origin impacts the gut microbiota (Fujiwara et al., 2008; Vilson et al., 2018). This may relate not only to parent genetics but also to the host of maternal factors that can affect the development of progeny GMB, including mode of delivery, maternal diet, maternal stress, and maternal antibiotic use (Bailey et al., 2004; Friswell et al., 2010; Stokholm et al., 2014; Golubeva et al., 2015; Walker et al., 2017; Zhang et al., 2021). For these reasons, if the study being examined utilises organisms from differing litters (from separate mothers or separate deliveries from the same mother) that have not yet reached their mature adult development and are not randomly assorted between research arms, it should be classified as having a high ROB. Suppose a study uses organisms from the same mother and litter or randomly assort progeny from different mothers and litters. In that case, it should be classified as having a low ROB.

Regarding sourcing of animal models, subjects sourced from different vendors have substantial differences in GMB at baseline (Rasmussen et al., 2019; Wolff et al., 2020; Long et al., 2021). The microbiological or physiological basis of these effects is unknown but may be due to differential exposures to environmental or infectious factors between vendors (Mandal et al., 2020).

Extreme diet

Dietary differences have been shown to alter the abundance of most gut microbes (Daniel et al., 2014; Do et al., 2018; Ang et al., 2020; Li et al., 2021). Because of this, maintaining the diet of interest is essential to avoid introducing confounding bias to the study. However, it may not always be possible to strictly control diet. This is especially relevant to clinical studies involving humans. In this situation, an evaluation of bias must note how a study documented these diet variations.

GMB normalisation

It is important to assure organisms being studied in research have similar baseline GMB. This allows for more definitive inference as to the effect of the intervention. Several strategies have been used to make the GMB as similar as possible over time. Removal of the entire GMB through the use of germ-free mice can allow for the artificial seeding of a select group of organisms (Yi and Li, 2012; Kennedy et al., 2018). However, the use of these mice necessarily limits the generalizability of a study. For this reason, research often uses organisms with populated GMBs and relies instead on antibiotics to homogenise the microbiome. The use of antibiotics introduces additional risks of bias which must be considered when evaluating a study (Theriot et al., 2016). The most significant ROB arises from beginning the intervention of interest before the gut microbiota has stabilised after normalisation with antibiotics. The GMB continues to fluctuate unpredictably for long periods following antibiotic administration (Merenstein et al., 2021). This variance has been found for at least a year after antibiotic

usage in humans and for times ranging between 1 week and 16 weeks in mice depending on the length of the course of antibiotics used (Rashid et al., 2015; Elvers et al., 2020; Zhu et al., 2021). However, short, or single doses of antibiotics such as those often used to normalise the microbiome allow for substantial stabilisation of the GMB within 7 days (Gu et al., 2020).

A third method used to standardise the GMB is to intermix the bedding of multiple cages and then redistribute it (Miyoshi et al., 2018). This method is less invasive than antibiotic usage and has a lower risk of long-term impact on the GMB than the use of antibiotics. The use of homogenisation of the bedding allows for similar microbiomes to develop in more mice than can be practically housed in a single cage, where the organisms also share all of their bedding (McCafferty et al., 2013).

Because of the impact of different methods of GMB normalisation, it is critical to note the method that was used to normalise the GMB and how long before the intervention this normalisation was completed.

Selection bias

Extreme genotype

Host genotype shows a stable and heritable impact on GMB composition (Goodrich et al., 2016). In the context of GMB research, extreme genotype selection refers to the selection of GMB subjects with genotypes that vary significantly between subjects within a study. Selection of subjects with identical or similar genetic makeup limits genotype confounding effects. A subject with an established history of use along with maximised genetic correlation can be considered a low risk of selection bias. For example, while inbred Balb/C mice do have an extreme genotype, they also have a long-established history of use in immune modulation studies with their known Th2 immune response wherein they exhibit low IFN γ and high IL-4 production (Mills et al., 2000; Watanabe et al., 2004; Khan et al., 2022). Furthermore, prior literature has established the correlation between subject genetics and variation in the GMB population and subsequent disease states (Xu et al., 2020).

Randomisation or demographic balancing sufficiently applied

Randomisation is essential in ensuring subject-level differences between participants in the intervention and control groups can be attributed to chance alone. It is a standard method that attempts to create the necessary pre-intervention equivalence between groups, allowing for conclusions based on the effect of the intervention. In trials where randomisation was not appropriately utilised, the outcome was overestimated by up to 40% compared to trials where randomisation was utilised (Suresh, 2011). If randomisation was not applied, implementing demographic balancing is an appropriate measure to ensure adequate control and intervention arms distribution. Any demographic balancing performed should be sufficiently described in the study. This method focuses on ensuring each group is demographically balanced at baseline to lessen the difference between groups and utilise randomisation if no subject background information is available (Saint, 2015). Both randomisation and demographic balancing can be applied to human and animal model studies. For example, in studies utilising syngeneic mice, randomisation must be performed outside the scope of human intervention in that random number generators should assign mice numbers which can then correlate to intervention and control groups, hence this places randomisation outside the scope of human influence, limiting bias to a maximum degree. In syngeneic animals, demographic balancing would have a limited impact on the bias, however, wherein studies utilise genetically unrelated animals, the need for implementation of both randomisation and demographic balancing is necessary for limiting substantial bias (Hirst et al., 2014). Similar principles apply in human studies. Given a majority of human studies utilise genetically unrelated subjects, randomisation is required to avoid the high ROB. In human studies, a step beyond randomisation should be taken, that is,

implementing blinded randomisation with a description of the randomisation protocol to give the reader the ability to discern breaks in randomisation or similar bias control methods within the study (Chalmers et al., 1983).

Classification of intervention

Intervention bias

Bias in intervention can occur when interventions or outcomes are inappropriately selected for or measured. In non-differential misclassification, test subjects' exposures are misidentified, and they are categorised into the wrong group (McCoy, 2017). This misclassification can dilute the effect of the intervention causing effect estimates to favour the null (LaMorfe, 2016). The probability of non-differential misclassification is equal across all groups. Bias may be reduced by ensuring a proper background check on test subjects and equalising any differences. On the other hand, differential misclassification occurs when the misclassification of exposure or outcome is not equal between subjects and is less easily predictable in whether it will bias results towards or away from the null. Therefore, the probability of assigning subjects to the wrong group differs based on the individual. This may also introduce recall bias towards recalling specific exposures because the subject has the disease state versus a subject that does not. In GMB studies, this may present in the form of researchers explaining results that show a significant effect as attributed to specific causes but leaving out explanations for non-significant results. Because this type of misclassification is more applicable in case studies, it is less relevant for animal studies but can be prominent in human studies (Spencer et al., 2018).

Validation of method

The establishment of an effective intervention is imperative for a successful study. Before the experiment, researchers must verify that their chosen intervention method will produce the intended effect. In studies where this is not done, the produced results may or may not be relied on because the protocol was never validated. Verification can be internal (tested and proved by the researchers) or external (via other established studies). If the study calls for a particular disease state to be expressed, it must be validated that the test subjects have the disease state. In studies that call for a specific procedure, there can be potential bias in how the readers know the procedure was correctly obtained if it is not reported. For example, in microbiome hypertension studies, animal subjects were tested based on blood pressure measurements by a well-established method, tail-cuff plethysmography (Marques et al., 2019). If a lesser-known and validated method was used, it could introduce a high ROB if researchers did not verify that their method was accurate. When testing for the effect of a disease state as influenced by the microbiome, it is helpful to transplant the experimental group microbiome into a germ-free animal model to confirm the effect. This reduces an intermediate ROB by demonstrating that the effect of the intervention is associated with the levels of change in the microbiome (Gottfredson et al., 2015).

Deviation from intervention

It is well understood that experiments that deviate from their initial protocol have an increased potential for bias in their study should they decide to include data prior to the deviation. Therefore, all deviations from the protocol should be well documented with time stamps, and the data included in the study should also include the time at which it was collected – either post-protocol or pre-protocol addendum. Rationale and limitations should also be included should researchers decide to include data from any time the protocol was different.

Missing data

Missing data is prevalent in many academic disciplines, from the social to biomedical sciences, and may contribute to bias in any given study. GMB research likewise suffers from inadequate consideration of missing data and the statistical methods to address it. To begin, two types of missing data should be distinguished: missing data due to patient drop-out in clinical, longitudinal studies and missing data as a result of inadequate sequencing depth leading to “false zeroes” in the microbiome genetic data. Both have the potential to increase ROB.

Cause/category of missing data

Missing data falls into multiple categories based on the mechanism of missingness: Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing Not at Random (MNAR) (Groenwold and Dekkers, 2020). These categories apply assumptions to missing data based on the cause. MCAR assumes that data is missing due to a factor entirely unrelated to the study. MAR assumes data is missing due to observed variables relevant to the study. MNAR assumes data is missing based on unknown or not quantifiable variables to the authors. MAR and MNAR are most relevant to clinical research, specifically in regard to patient drop-out, including clinical GMB trials (Pugh et al., 2021). Sampling zeroes in microbiome data are a more generalised form of missing data but are primarily reminiscent of MAR (Kaul et al., 2017a, 2017b). Each of these areas will be further discussed in the following sections. Under MAR, studies may utilise various statistical imputation techniques to replace missing data, though the most well-known and effective method is multiple imputations (Spineli et al., 2015). With MNAR, various statistical modelling techniques may address missing data. Such techniques are further discussed in relation to GMB studies in section “Sequencing depth and sampling zeroes.” The distinction between MAR and MNAR also indicates whether bias related to missing data is entirely removable in analysis – the former can, while the latter cannot (Mack et al., 2018). This should not be confused with the notion that MNAR assumptions immediately denote a study as biased. If the missingness in MNAR or MAR is independent of the outcome, then the study may be unbiased in regard to missing data. Thus, a study with MNAR data is not necessarily high ROB.

Notably, a significant number of studies do not clearly state the mechanism of missingness or adjust for missing data (Carpenter and Smuk, 2021). It is important that studies distinguish the mechanism of missingness or explain relevant missing data. If a study does not acknowledge missingness in data or ensures the absence of missing data, the study may be considered high ROB. If a study acknowledges missing data but does not adequately address it through MAR/MNAR distinction and proper statistical techniques related to its missing data category, then the study may be considered intermediate ROB. If a study demonstrates all of this, it may be considered low ROB.

Subject drop-out

Missing data in the form of patient drop-out has a marked effect on statistical power, type 1 error, and various outcome measures (Thompson et al., 2011; Fiero et al., 2016; Cai et al., 2020). In traditional clinical research, missing data has a clear effect on useful measures, such as relative risk and risk ratio calculations. Further, although researchers attempt to minimise drop-out and its statistical effects, drop-out ratios were reported to be greater than 40% depending on the study and the degree of unpleasantness in medical interventions to the patient (Schnicker et al., 2013; Li et al., 2021). Consequently, it has been proposed that a 20% drop-out ratio is reasonable (Furlan et al., 2009; Cramer et al., 2016). Interestingly, it has been shown that faecal sampling of patients in GMB studies has not been a significant reason for drop-out, suggesting typical sources of patient non-retention (Vandeputte et al., 2017). The effect of drop-out on statistical measures is expected to be the same in clinical GMB trials. Despite drop-out being common in clinical studies, its effect on

outcome measures involving microbial compositional data (e.g., beta diversity) is not currently well described in clinical GMB studies. However, it is expected that such measurements relying on consistent analysis from a wide array of samples will be biased if there is an inadequate sampling size.

The effect of bias comes into effect when there is an interpretation between samples, in that missing data prevents consistent interpretation of genetic data through a larger body of samples. For example, microbiome samples stratified by disease state versus control should be held to higher statistical power, similar to traditional clinical studies. Yet, the complexity of GMB genetic analysis often prevents large sample sizes from being a practical implementation due to costs unless utilising less-expensive protocols such as those involving qPCR to monitor microbial composition at high taxonomic levels (i.e., phyla) (Koliada et al., 2020). Some studies demonstrate shallow shotgun metagenomic sequencing as an alternative methodology for large, longitudinal GMB studies (Xu et al., 2021). Nonetheless, making interpretations in GMB data between samples stratified by host conditions may need to be more consistent and accurate when samples are unavailable from a patient drop-out. Based on the literature of other areas in clinical research as discussed, it is again reasonable to assert that drop-out will influence outcome measures if authors make interpretations across hosts of varying condition states.

Due to few clinical studies analysing the effect of drop-out on GMB outcomes, it is reasonable to use a 20% patient drop-out ratio, as many clinical trials traditionally utilise. GMB studies that have a high patient dropout are considered high ROB. GMB studies that have low patient drop-out are considered low ROB.

Sequencing depth and sampling zeroes

GMB researchers should consider sequencing depth as a contributor to missing data and subsequent bias. It is established that low-sequencing depth (2000 single-end reads per sample) can adequately predict the same diversity patterns as high-depth sequencing (on the scale of millions of reads per sample) (Caporaso et al., 2011; Lundin et al., 2012; Xiao et al., 2018). Experiments that quantify GMB outcome measures (like alpha and beta diversity) should utilise the same depth for all samples. Bias would be introduced if different sequencing depths are used for a set of samples. It should be noted, however, that false zeroes influence microbiome genetic data at both high and low depth. While true zeroes (or biological zeroes) represent true taxonomic absences, false zeroes (or sampling zeroes) represent a lack of sequencing depth to adequately detect certain microbial taxa. Notably, low sequencing depth, as is often the case of 16S rRNA sequencing, may not detect low abundance taxa or low taxa (subspecies) due to lower resolution. Though whole genome sequencing (WGS), such as shotgun metagenomic sequencing, utilises high sequencing depth to sequence entire genomes, sampling zeroes still persist (Pereira-Marques et al., 2019).

At the time of writing, this issue of zero-inflation – or the excess of sampling zeroes at high and low depth – and the resulting bias in GMB genetic data is an active area of research. Interestingly, relatively few studies utilise any statistical modelling to correct for such missing data. Yet, various modelling techniques were recently developed to address zero-inflation (Ha et al., 2020; Zhang et al., 2020; Deek and Li, 2021). Similar to modelling techniques, imputation is a method traditionally used to address missing data in the form of patient drop out, but a promising imputation method is recently available to also deal with GMB sampling zeroes. Previous studies showed an increase in Pearson correlation from 0.59 (between 16S and WGS in non-corrected data) to 0.64 (between 16S and WGS in corrected data) (Jiang et al., 2021). There were also marked differences in mean and standard deviation of abundances per taxon between corrected and non-corrected data. This suggests greater homogeneity of samples across sequencing methods if imputation is utilised to correct data. However, as our article focuses on the role of bias in GMB research, we do not yet place best-practice recommendations for a particular method of missing data correction.

As of date, few GMB studies utilise statistical techniques to correct for sampling zeroes. Furthermore, common bioinformatics pipelines (such as QIIME2) do not incorporate such techniques into data-correction programs.

As such, the available literature suggests future GMB studies that do not consider sampling zeroes and lack a statistical technique for missing data correction may be considered high ROB. Studies that utilise missing data correction may be considered low ROB. These data correction methods, once more, include various modelling techniques or imputation.

Measurement of outcomes

Sample collection

Currently, there is no standard method for sample collection for GMB studies. While biopsy of the lower intestine provides a controlled sampling site and an accurate microbiota account, it is expensive, time-consuming, and unsuitable for healthy control groups. In contrast, the faecal collection is non-invasive and cost-effective (Tang et al., 2020). Thus, it is a standard sampling method in both clinical and research applications. However, faecal sample collection introduces temporal inconsistency which can be a source of bias.

Faecal samples collected at different times of the day are at risk for inaccurate representation of the absolute abundance of gut microbiota (Caporaso et al., 2011). Specifically for mouse studies, the snapshots of the microbiota provided by the faecal samples are more accurate and consistent within treatment groups when collected in the morning due to the nocturnal feeding nature of mice (Jones et al., 2021). For studies involving subjects with unpredictable and inconsistent bowel movements, samples should be preserved immediately after defecation as oxidation of the outer layer can alter the microbiota (Pepper and Rosenfeld, 2012). Specifically, *Firmicutes* and *Bifidobacteria* spp. are two known phylum that are unstable in the outer microenvironment when exposed to oxygen (Gorzalak et al., 2015). Therefore, to minimise the differential errors, the methods of measurement must be consistent between control and intervention groups.

Blinding

In a GMB study, the primary outcome is based on definitive and objective genetic sequencing. Therefore, assessor bias is typically negligible, and a low ROB is expected (Higgins et al., 2022).

Reporting of results

Selection of reported results

Selective reporting of results can lead to biased interpretations of significance and or non-significance via particular selection of results from multiple outcome measures in estimating outcome effect. Bias in the selection of reported results can be difficult to detect without access to a protocol from which one can compare pre-specified intended outcomes of interest to the outcomes analysed in the published paper (Heneghan et al., 2019). Often, results are selected for significance, omitted for non-significance, or omitted for adverse effect of intervention (Dwan et al., 2013; Hedin et al., 2016; Van der Steen et al., 2019).

Validation test

Four medical students with no prior experience in ROB assessment were recruited to test this tool by using it to independently assess ROB on three selected studies of similar length in a predetermined

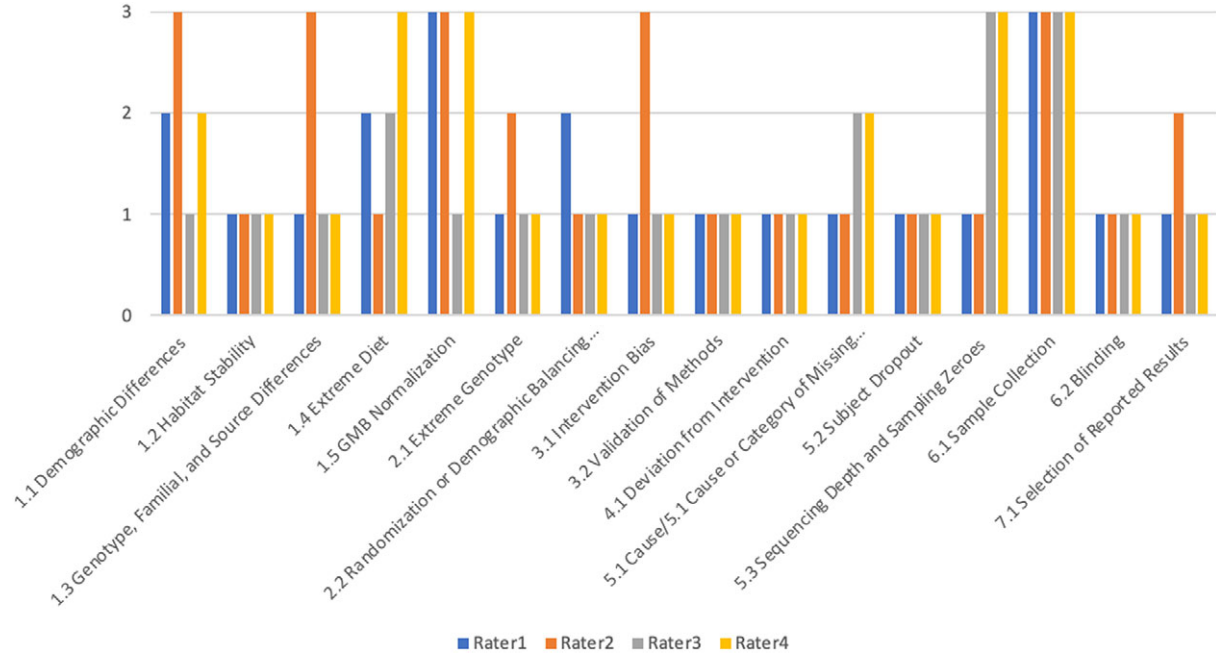


Figure 1. Inter-rater variability in ROB determinations by subdomain for validation test study 1 by Wu et al. (2017), where “1” on the y-axis indicates that the rater determined the study to be at low ROB for the subdomain indicated on the x-axis; “2” indicates medium ROB and “3” indicates a high ROB determination by the individual rater.

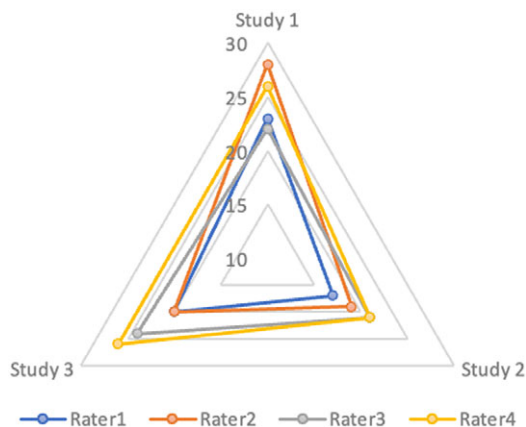


Figure 2. Inter-rater variability in ROB determinations by subdomain for validation test on study 2 by Mohammed et al. (2020), where “1” on the y-axis indicates that the rater determined the study to be at low ROB for the subdomain indicated on the x-axis; “2” indicates medium ROB and “3” indicates a high ROB determination by the individual rater.

sequence (Wu et al., 2017; Mohammed et al., 2020; Saunders et al., 2020). Subjects were provided with the manuscript and ROB rubric. They were asked to track time to completion per study and complete the ROB rubric for each study. Subjects assessed ROB in an average of 44.75 minutes per study with time to completion generally decreasing from the first study assessed to the last study assessed.

Inter-rater variability was assessed by assigning values of 1, 2, and 3 to low, medium, and high ROB in order to construct visual representations of rater scores in each sub-domain of bias and to compare summed ROB scores between raters for each study. Figures 1–3 demonstrate variability within a study in each subdomain of bias assessed by this tool between raters. The figures demonstrate similar ROB judgements between at least three of four raters in the majority of subdomains across the three studies assessed.

Figure 4 demonstrates variation in summed ROB score by rater for each of the three studies. It shows the decreasing magnitude of difference between raters’ summed ROB scores with each subsequent use of the tool from a max-score min-score difference of six points in study 1 and study 3, and of four points in study 2 out of 45 possible points. One-way ANOVA test of rater subdomain scores across all subdomains for each study returned *p*-values of 0.554, 0.568, and 0.399 for study 1, study 2, and study 3, respectively indicating no significant difference between overall ROB assessment scores between raters of the same study. First-time ROB assessors using this tool showed a relatively high degree of concordance in ROB determination at the subdomain level and in the magnitude of summed ROB score.

Conclusion

ROB assessment is a crucial step in systematic review and meta-analysis to assess the quality of information being collected. By outlining common sources of bias that can impact GMB research following the structure of the ROBINS-I tool, this tool can serve as an adjunct to improve and standardise ROB assessment of GMB studies. A standardised ROB assessment for GMB studies will improve the accuracy of risk assessment, improve reproducibility between researchers, and promote the inclusion of high-quality information in systematic reviews and meta-analyses of the GMB.

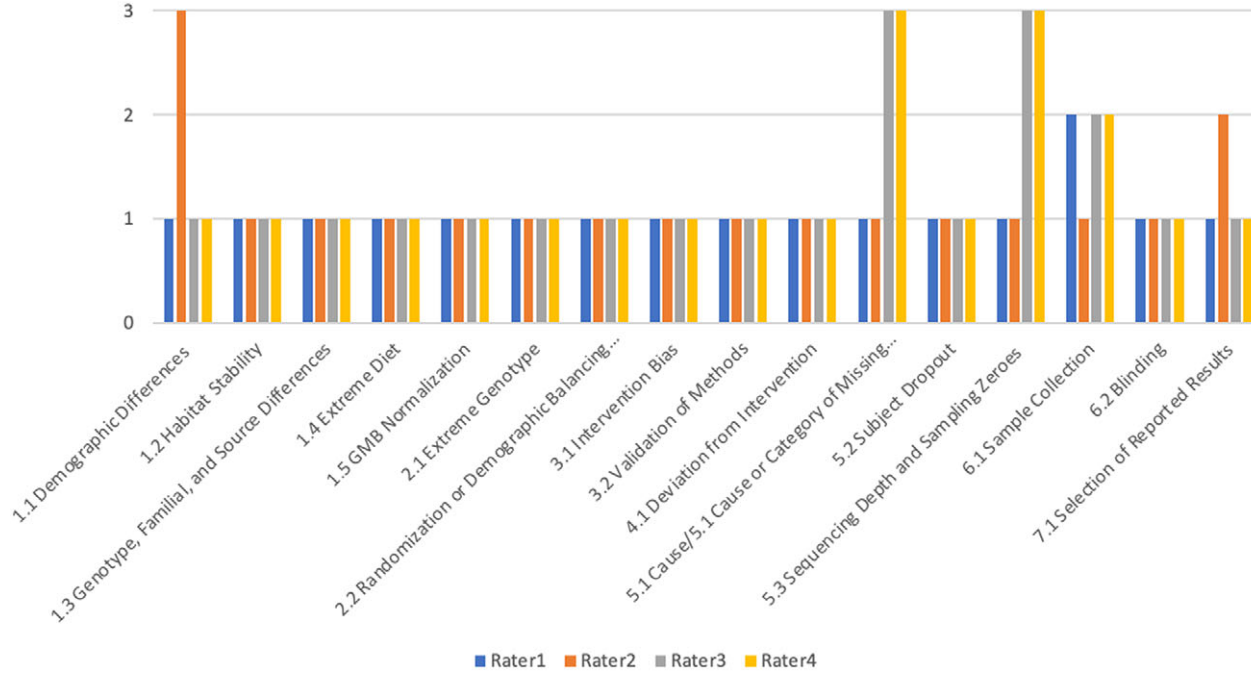


Figure 3. Inter-rater variability in ROB determinations by subdomain for validation test on study 3 by Saunders et al. (2020), where “1” on the y-axis indicates that the rater determined the study to be at low ROB for the subdomain indicated on the x-axis; “2” indicates medium ROB and “3” indicates a high ROB determination by the individual rater.

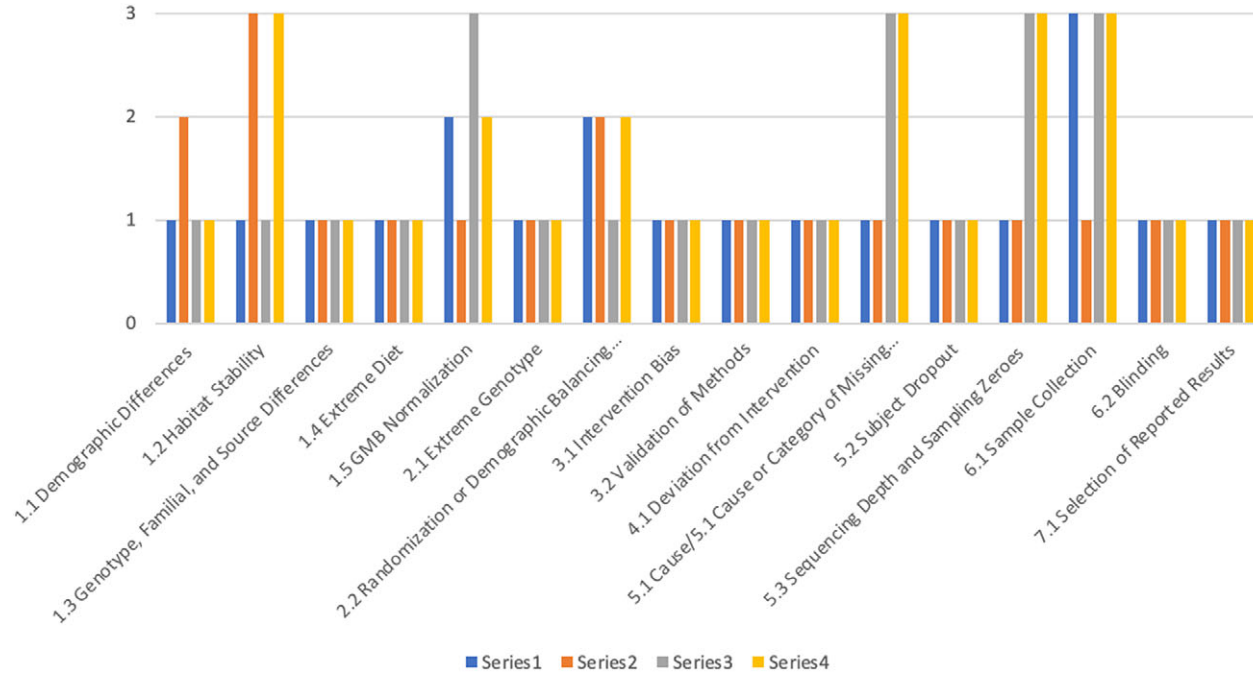


Figure 4. Visual representation comparing summed ROB score (as determined by assigning point values of 1, 2, and 3 to low, medium, and high ROB respectively) by rater for each of the three studies assessed in the validation test where each increasingly large concentric triangle indicates an increase of 5 points.

Data availability statement. Following the journal's policy for supporting research transparency and reproducibility, we will make all data and protocols available to readers.

Author contribution. M.R., L.-L.Y., and T.L. conceived and planned the concept. A.M. consulted on methods in meta-analysis and assisted in structuring the approach. T.L. and M.R. drafted the subdomains of interest with guidance from A.M. T.L., C.L., T.T.T., A.S.M., H.Y.L., A.O., and D.M. researched influencers on bias subdomains. T.L., C.L., T.T.T., A.S.M., H.Y.L., A.O., A.K., and D.M. wrote the body of manuscript describing subdomains of bias. T.L., C.L., T.T.T., A.S.M., H.Y.L., A.O., and D.M. described summary criteria contained in Table 1. L.-L.Y., T.L., and M.R. edited the manuscript with input from all authors. M.O. consulted in manuscript editing. L.-L.Y. guided manuscript formatting and editing. T.L. planned and performed validation test. M.R., L.-L.Y., and M.O. consulted on and reviewed validation test design, methods, and interpretation of results. T.L. created figures and table. L.-L.Y. guided all phases of development. All authors made valuable intellectual contributions in discussion and creation of this manuscript.

Disclosure statement. The authors declare no conflicts of interest.

Funding. This work was supported by the Iowa Osteopathic Education and Research (IOER) Foundation.

References

- Ang QY, Alexander M, Newman JC, Tian Y, Cai J, Upadhyay V, Turnbaugh JA, Verdin E, Hall KD, Leibel RL, Ravussin E, Rosenbaum M, Patterson AD and Turnbaugh PJ (2020) Ketogenic diets alter the gut microbiome resulting in decreased intestinal Th17 cells. *Cell* **181**(6), 1263–1275.e16. <https://doi.org/10.1016/j.cell.2020.04.027>
- Bailey MT, Lubach GR and Coe CL (2004) Prenatal stress alters bacterial colonization of the gut in infant monkeys. *Journal of Pediatric Gastroenterology and Nutrition* **38**(4), 414–421. <https://doi.org/10.1097/00005176-200404000-00009>
- Cai X, Gewandter JS, He H, Turk DC, Dworkin RH and McDermott MP (2020) Estimands and missing data in clinical trials of chronic pain treatments: Advances in design and analysis. *Pain* **161**(10), 2308–2320. <https://doi.org/10.1097/j.pain.0000000000001937>
- Campbell JH, Foster CM, Vishnivetskaya T, Campbell AG, Yang ZK, Wymore A, Palumbo AV, Chesler EJ and Podar M (2012) Host genetic and environmental effects on mouse intestinal microbiota. *The ISME Journal* **6**(11), 2033–2044. <https://doi.org/10.1038/ismej.2012.54>
- Capdevila S, Giral M, Ruiz de la Torre JL, Russell RJ and Kramer K (2007) Acclimatization of rats after ground transportation to a new animal facility. *Laboratory Animals* **41**(2), 255–261. <https://doi.org/10.1258/002367707780378096>
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N and Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* **108**(Suppl 1), 4516–4522. <https://doi.org/10.1073/pnas.1000080107>
- Carpenter JR and Smuk M (2021) Missing data: A statistical framework for practice. *Biometrical journal. Biometrische Zeitschrift* **63**(5), 915–947. <https://doi.org/10.1002/bimj.202000196>
- Chalmers TC, Celano P, Sacks HS and Smith H (1983) Bias in treatment assignment in controlled clinical trials. *The New England Journal of Medicine* **309**(22), 1358–1361. <https://doi.org/10.1056/NEJM198312013092204>
- Cramer H, Haller H, Dobos G and Lauche R (2016) A systematic review and meta-analysis estimating the expected dropout rates in randomized controlled trials on yoga interventions. *Evidence-Based Complementary and Alternative Medicine: eCAM* **2016**, 5859729. <https://doi.org/10.1155/2016/5859729>
- Daniel H, Gholami AM, Berry D, Desmarchelier C, Hahne H, Loh G, Mondot S, Lepage P, Rothballer M, Walker A, Böhm C, Wenning M, Wagner M, Blaut M, Schmitt-Kopplin P, Kuster B, Haller D and Clavel T (2014) High-fat diet alters gut microbiota physiology in mice. *The ISME Journal* **8**(2), 295–308. <https://doi.org/10.1038/ismej.2013.155>
- Deek RA and Li H (2021) A zero-inflated latent Dirichlet allocation model for microbiome studies. *Frontiers in Genetics* **11**, 602594. <https://doi.org/10.3389/fgene.2020.602594>
- Do MH, Lee E, Oh MJ, Kim Y and Park HY (2018) High-glucose or -fructose diet cause changes of the gut microbiota and metabolic disorders in mice without body weight change. *Nutrients* **10**(6), 761. <https://doi.org/10.3390/nu10060761>
- Dwan K, Gamble C, Williamson PR and Kirkham JJ (2013) Systematic review of the empirical evidence of study publication bias and outcome reporting bias — an updated review. *PLoS One* **8**(7), e66844. <https://doi.org/10.1371/journal.pone.0066844>
- Evers KT, Wilson VJ, Hammond A, Duncan L, Huntley AL, Hay AD and van der Werf ET (2020) Antibiotic-induced changes in the human gut microbiota for the most commonly prescribed antibiotics in primary care in the UK: A systematic review. *BMJ Open* **10**(9), e035677. <https://doi.org/10.1136/bmjopen-2019-035677>
- Fiero MH, Huang S, Oren E and Bell ML (2016) Statistical analysis and handling of missing data in cluster randomized trials: A systematic review. *Trials* **17**, 72. <https://doi.org/10.1186/s13063-016-1201-z>
- Friswell MK, Gika H, Stratford IJ, Theodoridis G, Telfer B, Wilson ID and McBain AJ (2010) Site and strain-specific variation in gut microbiota profiles and metabolism in experimental mice. *PLoS One* **5**(1), 8584. <https://doi.org/10.1371/journal.pone.0008584>

- Fujiwara R, Watanabe J and Sonoyama K (2008) Assessing changes in composition of intestinal microbiota in neonatal BALB/c mice through cluster analysis of molecular markers. *The British Journal of Nutrition* **99**(6), 1174–1177. <https://doi.org/10.1017/S0007114507862349>
- Furlan AD, Pennick V, Bombardier C, van Tulder M, Board E and Group CBR (2009) 2009 updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine* **34**(18), 1929–1941. <https://doi.org/10.1097/BRS.0b013e3181b1c99f>
- Golubeva AV, Crampton S, Desbonnet L, Edge D, O’Sullivan O, Lomasney KW, Zhdanov AV, Crispie F, Moloney RD, Borre YE, Cotter PD, Hyland NP, O’Halloran KD, Dinan TG, O’Keefe GW and Cryan JF (2015) Prenatal stress-induced alterations in major physiological systems correlate with gut microbiota composition in adulthood. *Psychoneuroendocrinology* **60**, 58–74. <https://doi.org/10.1016/j.psyneuen.2015.06.002>
- Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, Spector TD, Bell JT, Clark AG and Ley RE (2016) Genetic determinants of the gut microbiome in UK twins. *Cell Host & Microbe* **19**(5), 731–743. <https://doi.org/10.1016/j.chom.2016.04.017>
- Gozelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M and Gibson DL (2015) Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PLoS One* **10**(8), e0134802. <https://doi.org/10.1371/journal.pone.0134802>
- Gottfredson DC, Cook TD, Gardner FE, Gorman-Smith D, Howe GW, Sandler IN and Zafft KM (2015) Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science* **16**(7), 893–926. <http://doi.org/10.1007/s11121-015-0555-x>
- Groenwold R and Dekkers OM (2020) Missing data: The impact of what is not there. *European Journal of Endocrinology* **183**(4), E7–E9. <https://doi.org/10.1530/EJE-20-0732>
- Gu SL, Gong Y, Zhang J, Chen Y, Wu Z, Xu Q, Fang Y, Wang J and Tang LL (2020) Effect of the short-term use of florquinolone and β -lactam antibiotics on mouse gut microbiota. *Infection and Drug Resistance* **13**, 4547–4558. <https://doi.org/10.2147/IDR.S281274>
- Ha MJ, Kim J, Galloway-Peña J, Do KA and Peterson CB (2020) Compositional zero-inflated network estimation for microbiome data. *BMC Bioinformatics* **21**(Suppl 21), 581. <https://doi.org/10.1186/s12859-020-03911-w>
- Heneghan T (2019, March 5) Outcome reporting bias. Catalog of Bias. Available at <https://catalogofbias.org/biases/outcome-reporting-bias/> (accessed 6 September 2022).
- Hedin R, Umerham B, Detweiler B, Kollmorgen L and Vassar M (2016) Publication bias and nonreporting found in majority of systematic reviews and meta-analyses in anesthesiology journals. *Anesthesia & Analgesia* **123**(4), 1018–1025. <https://doi.org/10.1213/ANE.0000000000001452>
- Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ and Welch VA (eds) (2022), *Cochrane Handbook for Systematic Reviews of Interventions version 6.3* (updated February 2022). Cochrane. Available at www.training.cochrane.org/handbook.
- Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C and Heneghan C (2014) The need for randomization in animal trials: An overview of systematic reviews. *PLoS One* **9**(6), e98856. <https://doi.org/10.1371/journal.pone.0098856>
- Hoy YE, Bik EM, Lawley TD, Holmes SP, Monack DM, Theriot JA and Relman DA (2015) Variation in taxonomic composition of the fecal microbiota in an inbred mouse strain across individuals and time. *PLoS One* **10**(11), e0142825. <https://doi.org/10.1371/journal.pone.0142825>
- Hufeldt MR, Nielsen DS, Vogensen FK, Midtvedt T and Hansen AK (2010) Family relationship of female breeders reduce the systematic inter-individual variation in the gut microbiota of inbred laboratory mice. *Laboratory Animals* **44**(4), 283–289. <https://doi.org/10.1258/la.2010.010058>
- Jiang R, Li WV and Li JJ (2021) mbImpute: An accurate and robust imputation method for microbiome data. *Genome Biology* **22**(1), 192. <https://doi.org/10.1186/s13059-021-02400-4>
- Jones J, Reinke SN, Ali A, Palmer DJ and Christophersen CT (2021) Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations. *Scientific Reports* **11**(1), 13964. <https://doi.org/10.1038/s41598-021-93031-z>
- Kaul A, Davidov O and Peddada SD (2017a) Structural zeros in high-dimensional data with applications to microbiome studies. *Biostatistics (Oxford, England)* **18**(3), 422–433. <https://doi.org/10.1093/biostatistics/kxw053>
- Kaul A, Mandal S, Davidov O and Peddada SD (2017b) Analysis of microbiome data in the presence of excess zeros. *Frontiers in Microbiology* **8**, 2114. <https://doi.org/10.3389/fmicb.2017.02114>
- Kennedy EA, King KY and Baldrige MT (2018) Mouse microbiota models: Comparing germ-free mice and antibiotics treatment as tools for modifying gut bacteria. *Frontiers in Physiology* **9**, 1534. <https://doi.org/10.3389/fphys.2018.01534>
- Khan AI, Rehman AU, Farooqui NA, Siddiqui NZ, Ayub Q, Ramzan MN, Wang L and Xin Y (2022) Effects of shrimp peptide hydrolysate on intestinal microbiota restoration and immune modulation in cyclophosphamide-treated mice. *Molecules* **27**(5), 1720. <https://doi.org/10.3390/molecules27051720>
- Kim YS, Unno T, Kim BY and Park MS (2020) Sex differences in gut microbiota. *The World Journal of Men’s Health* **38**(1), 48–60. <https://doi.org/10.5534/wjmh.190009>

- Koliada A, Moseiko V, Romanenko M, Piven L, Lushchak O, Kryzhanovska N, Guryanov V and Vaiserman A** (2020) Seasonal variation in gut microbiota composition: Cross-sectional evidence from Ukrainian population. *BMC Microbiology* **20**(1), 100. <https://doi.org/10.1186/s12866-020-01786-8>
- LaMorfe W** (2016) Information Bias (observation bias), June 8. Available at https://sphweb.bumc.bu.edu/otlt/mph-modules/ep/ep713_bias/EP713_Bias4.html (accessed 26 March 2022).
- Laukens D, Brinkman BM, Raes J, De Vos M and Vandenaabeele P** (2016) Heterogeneity of the gut microbiome in mice: Guidelines for optimizing experimental design. *FEMS Microbiology Reviews* **40**(1), 117–132. <https://doi.org/10.1093/femsre/fuv036>
- Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K, Sinha R, Peterson DA, Walter J, Benson AK and Pomp D** (2014) Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biology* **15**(12), 552. <https://doi.org/10.1186/s13059-014-0552-6>
- Li L, Pei H, Liu Z and Zhang J** (2021) Analysis of risk factors and construction of prediction model of drop out from peritoneal dialysis. *Medicine* **100**(3), e24195. <https://doi.org/10.1097/MD.00000000000024195>
- Li Y, Yang X, Zhang J, Jiang T, Zhang Z, Wang Z, Gong M, Zhao L and Zhang C** (2021) Ketogenic diets induced glucose intolerance and lipid accumulation in mice with alterations in gut microbiota and metabolites. *MBio* **12**(2), e03601–e03620. <https://doi.org/10.1128/mBio.03601-20>
- Lipinski JH, Zhou X, Gurczynski SJ, Erb-Downward JR, Dickson RP, Huffnagle GB, Moore BB and O'Dwyer DN** (2021) Cage environment regulates gut microbiota independent of toll-like receptors. *Infection and Immunity* **89**(9), e0018721. <https://doi.org/10.1128/IAI.00187-21>
- Liu A, Lv H, Wang H, Yang H, Li Y and Qian J** (2020) Aging increases the severity of colitis and the related changes to the gut barrier and gut microbiota in humans and mice. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* **75**(7), 1284–1292.
- Long LL, Svenson KL, Mourino AJ, Michaud M, Fahey JR, Waterman L, Vandegrift KL and Adams MD** (2021) Shared and distinctive features of the gut microbiome of C57BL/6 mice from different vendors and production sites, and in response to a new vivarium. *Lab Animal* **50**(7), 185–195. <https://doi.org/10.1038/s41684-021-00777-0>
- Lundin D, Severin I, Logue JB, Ostman O, Andersson AF and Lindström ES** (2012) Which sequencing depth is sufficient to describe patterns in bacterial α - and β -diversity? *Environmental Microbiology Reports* **4**(3), 367–372. <https://doi.org/10.1111/j.1758-2229.2012.00345.x>
- Mack C, Su Z and Westreich D** (2018) *Managing Missing Data in Patient Registries: Addendum to Registries for Evaluating Patient Outcomes: A User's Guide*, 3rd Edn. Rockville, MD: Agency for Healthcare Research and Quality (US). Types of Missing Data. Available at <https://www.ncbi.nlm.nih.gov/books/NBK493614/> (accessed 09 January 2022).
- Mandal RK, Denny JE, Waide ML, Li Q, Bhutiani N, Anderson CD, Baby BV, Jala VR, Egilmez NK and Schmidt NW** (2020) Temporospatial shifts within commercial laboratory mouse gut microbiota impact experimental reproducibility. *BMC Biology* **18**(1), 83. <https://doi.org/10.1186/s12915-020-00810-7>
- Marques FZ, Jama HA, Tsyganov K, Gill PA, Rhys-Jones D, Muralitharan RR, Muir J, Holmes A and Mackay CR** (2019) Guidelines for transparency on gut microbiome studies in essential and experimental hypertension. *Hypertension* **74**(6), 1279–1293. <https://doi.org/10.1161/hypertensionaha.119.13079>
- McCafferty J, Mühlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, Jobin C and Fodor AA** (2013) Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *The ISME Journal* **7**(11), 2116–2125. <https://doi.org/10.1038/ismej.2013.106>
- McCoy CE** (2017) Understanding the intention-to-treat principle in randomized controlled trials. *The Western Journal of Emergency Medicine* **18**(6), 1075–1078. <https://doi.org/10.5811/westjem.2017.8.35985>
- McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, Bastiaansen JW, Wang X, Kachman SD, Auwerx J, Williams RW, Benson AK, Peterson DA and Ciobanu DC** (2012) Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One* **7**(6), e39191. <https://doi.org/10.1371/journal.pone.0039191>
- Merenstein D, Fraser CM, Roberts RF, Liu T, Grant-Beurmann S, Tan TP, Smith KH, Cronin T, Martin OA, Sanders ME, Lucan SC and Kane MA** (2021) *Bifidobacterium animalis* subsp. *lactis* BB-12 protects against antibiotic-induced functional and compositional changes in human Fecal microbiome. *Nutrients* **13**(8), 2814. <https://doi.org/10.3390/nu13082814>
- Mills CD, Kincaid K, Alt JM, Heilman MJ and Hill AM** (2000) M-1/M-2 macrophages and the Th1/Th2 paradigm. *Journal of Immunology* **164**(12), 6166–6173. <https://doi.org/10.4049/jimmunol.164.12.6166>
- Miyoshi J, Leone V, Nobutani K, Musch MW, Martinez-Guryn K, Wang Y, Miyoshi S, Bobe AM, Eren AM and Chang EB** (2018) Minimizing confounders and increasing data quality in murine models for studies of the gut microbiome. *PeerJ* **6**, e5166. <https://doi.org/10.7717/peerj.5166>
- Mohammed A, Alghetaa HK, Zhou J, Chatterjee S, Nagarkatti P and Nagarkatti M** (2020) Protective effects of Δ^9 -tetrahydrocannabinol against enterotoxin-induced acute respiratory distress syndrome are mediated by modulation of microbiota. *British Journal of Pharmacology* **177**(22), 5078–5095.
- Montonye DR, Ericsson AC, Busi SB, Lutz C, Wardwell K and Franklin CL** (2018) Acclimation and institutionalization of the mouse microbiota following transportation. *Frontiers in Microbiology* **9**, 1085. <https://doi.org/10.3389/fmicb.2018.01085>

- Org E, Mehrabian M, Parks BW, Shipkova P, Liu X, Drake TA and Lusic AJ** (2016) Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes* **7**(4), 313–322.
- Pepper JW and Rosenfeld S** (2012) The emerging medical ecology of the human gut microbiome. *Trends in Ecology & Evolution* **27**(7), 381–384. <https://doi.org/10.1016/j.tree.2012.03.002>
- Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn LJ, Knetsch CW and Figueiredo C** (2019) Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Frontiers in Microbiology* **10**, 1277. <https://doi.org/10.3389/fmicb.2019.01277>
- Pugh SL, Brown PD and Enserro D** (2021) Missing repeated measures data in clinical trials. *Neuro-Oncology Practice* **9**(1), 35–42. <https://doi.org/10.1093/nop/npab043>
- Rashid MU, Zaura E, Buijs MJ, Keijser BJ, Crielaard W, Nord CE and Weintraub A** (2015) Determining the long-term effect of antibiotic administration on the human normal intestinal microbiota using culture and pyrosequencing methods. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* **60**(Suppl 2), S77–S84. <https://doi.org/10.1093/cid/civ137>
- Rasmussen TS, de Vries L, Kot W, Hansen LH, Castro-Mejía JL, Vogensen FK, Hansen AK and Nielsen DS** (2019) Mouse vendor influence on the bacterial and viral gut composition exceeds the effect of diet. *Viruses* **11**(5), 435. <https://doi.org/10.3390/v11050435>
- Saint U** (2015) Randomization does not help much, comparability does. *PLoS One* **10**(7), e0132102. <https://doi.org/10.1371/journal.pone.0132102>
- Saunders JM, Moreno JL, Ibi D, Sikaroodi M, Kang DJ, Muñoz-Moreno R, Dalmat SS, García-Sastre A, Gillevet PM, Dozmorov MG, Bajaj JS and González-Maeso J** (2020) Gut microbiota manipulation during the prepubertal period shapes behavioral abnormalities in a mouse neurodevelopmental disorder model. *Scientific Reports* **10**(1), 4697.
- Schnicker K, Hiller W and Legenbauer T** (2013) Drop-out and treatment outcome of outpatient cognitive-behavioral therapy for anorexia nervosa and bulimia nervosa. *Comprehensive Psychiatry* **54**(7), 812–823. <https://doi.org/10.1016/j.comppsych.2013.02.007>
- Singh G, Brass A, Cruickshank SM and Knight CG** (2021) Cage and maternal effects on the bacterial communities of the murine gut. *Scientific Reports* **11**(1), 9841. <https://doi.org/10.1038/s41598-021-89185-5>
- Spencer EA, Mahtani KR, Brassey J and Heneghan C** (2018) Misclassification bias. *Catalogue of bias 2018*. Available at <http://www.catalogueofbiases.org/biases/misclassificationbias> Accessed 09 January 2022.
- Spineli LM, Fleming PS and Pandis N** (2015) Addressing missing participant outcome data in dental clinical trials. *Journal of Dentistry* **43**(6), 605–618. <https://doi.org/10.1016/j.jdent.2015.03.007>
- Stokholm J, Schjørring S, Eskildsen CE, Pedersen L, Bischoff AL, Følsgaard N, Carson CG, Chawes BLK, Bønnelykke K, Mølgaard A, Jacobsson B, Krogfelt KA and Bisgaard H** (2014) Antibiotic use during pregnancy alters the commensal vaginal microbiota. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* **20**(7), 629–635. <https://doi.org/10.1111/1469-0691.12411>
- Suresh K** (2011) An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences* **4**(1), 8–11. <https://doi.org/10.4103/0974-1208.82352>
- Tang Q, Jin G, Wang G, Liu T, Liu X, Wang B and Cao H** (2020) Current sampling methods for gut microbiota: A call for more precise devices. *Frontiers in Cellular and Infection Microbiology* **10**, 151. <https://doi.org/10.3389/fcimb.2020.00151>
- Theriot CM, Bowman AA and Young VB** (2016) Antibiotic-induced alterations of the gut microbiota alter secondary bile acid production and allow for Clostridium difficile spore germination and outgrowth in the large intestine. *mSphere* **1**(1), e00045. <https://doi.org/10.1128/mSphere.00045-15>
- Thompson J, Berk M, Dean O, Kohlmann K, Jeavons S, Bush A and Copolov D** (2011) Who's left? Symptoms of schizophrenia that predict clinical trial dropout. *Human Psychopharmacology* **26**(8), 609–613. <https://doi.org/10.1002/hup.1253>
- Ticinesi A, Nouvenne A, Cerundolo N, Catania P, Prati B, Tana C and Meschi T** (2019) Gut microbiota, muscle mass and function in aging: A focus on physical frailty and sarcopenia. *Nutrients* **11**(7), 1633.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R and Gordon JI** (2008) A core gut microbiome in obese and lean twins. *Nature* **457**(7228), 480–484. <https://doi.org/10.1038/nature07540>
- Valles-Colomer M, Bacigalupe R, Vieira-Silva S, Suzuki S, Darzi Y, Tito RY, Yamada T, Segata N, Raes J and Falony G** (2021) Variation and transmission of the human gut microbiota across multiple familial generations. *Nature Microbiology* **7**(1), 87–96. <https://doi.org/10.1038/s41564-021-01021-8>
- Van der Steen JT, Ter Riet G, van den Bogert CA and Bouter LM** (2019) Causes of reporting bias: a theoretical framework. *F1000Research* **8**, 280. <https://doi.org/10.12688/f1000research.18310.2>
- Vandeputte D, Tito RY, Vanleuwen R, Falony G and Raes J** (2017) Practical considerations for large-scale gut microbiome studies. *FEMS Microbiology Reviews* **41**(Supplement_1), S154–S167. <https://doi.org/10.1093/femsre/fux027>
- Wilson Å, Ramadan Z, Li Q, Hedhammar Å, Reynolds A, Spears J, Labuda J, Pelker R, Björkstén B, Dicksved J and Hansson-Hamlin H** (2018) Disentangling factors that shape the gut microbiota in German shepherd dogs. *PLoS One* **13**(3), e0193507. <https://doi.org/10.1371/journal.pone.0193507>

- Walker RW, Clemente JC, Peter I and Loos RJF (2017) The prenatal gut microbiome: Are we colonized with bacteria in utero?: Colonization of the gut microbiome in utero. *Pediatric Obesity* **12**, 3–17. <https://doi.org/10.1111/ijpo.12217>
- Watanabe H, Numata K, Ito T, Takagi K and Matsukawa A (2004) Innate immune response in Th1- and Th2-dominant mouse strains. *Shock* **22**(5), 460–466. <https://doi.org/10.1097/01.shk.0000142249.08135.e9>
- Wolff NS, Jacobs MC, Haak BW, Roelofs J, de Vos AF, Hugenholtz F and Wiersinga WJ (2020) Vendor effects on murine gut microbiota and its influence on lipopolysaccharide-induced lung inflammation and gram-negative pneumonia. *Intensive Care Medicine Experimental* **8**(1), 47. <https://doi.org/10.1186/s40635-020-00336-w>
- Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, Ståhlman M, Olsson LM, Serino M, Planas-Félix M, Xifra G, Mercader JM, Torrents D, Burcelin R, Ricart W, Perkins R, Fernández-Real JM and Bäckhed F (2017) Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nature Medicine* **23**(7), 850–858.
- Xiao F, Yu Y, Li J, Juneau P and Yan Q (2018) Necessary sequencing depth and clustering method to obtain relatively stable diversity patterns in studying fish gut microbiota. *Current Microbiology* **75**(9), 1240–1246. <https://doi.org/10.1007/s00284-018-1516-y>
- Xu W, Chen T, Pei Y, Guo H, Li Z, Yang Y, Zhang F, Yu J, Li X, Yang Y, Zhao B and Wu C (2021) Characterization of shallow whole-metagenome shotgun sequencing as a high-accuracy and low-cost method by complicated mock microbiomes. *Frontiers in Microbiology* **12**, 678319. <https://doi.org/10.3389/fmicb.2021.678319>
- Xu F, Fu Y, Sun TY, Jiang Z, Miao Z, Shuai M, Gou W, Ling CW, Yang J, Wang J, Chen YM and Zheng JS (2020) The interplay between host genetics and the gut microbiome reveals common and distinct microbiome features for complex human diseases. *Microbiome* **8**(1), 145. <https://doi.org/10.1186/s40168-020-00923-9>
- Yi P and Li L (2012) The germfree murine animal: An important animal model for research on the relationship between gut microbiota and the host. *Veterinary Microbiology* **157**(1–2), 1–7. <https://doi.org/10.1016/j.vetmic.2011.10.024>
- Yoon CH, Ryu JS, Moon J and Kim MK (2021) Association between aging-dependent gut microbiome dysbiosis and dry eye severity in C57BL/6 male mouse model: A pilot study. *BMC Microbiology* **21**(1), 106.
- Zhang X, Guo B and Yi N (2020) Zero-inflated gaussian mixed models for analyzing longitudinal microbiome data. *PLoS One* **15**(11), e0242073. <https://doi.org/10.1371/journal.pone.0242073>
- Zhang Z, Li N, Chen R, Lee T, Gao Y, Yuan Z, Nie Y and Sun T (2021) Prenatal stress leads to deficits in brain development, mood related behaviors and gut microbiota in offspring. *Neurobiology of Stress* **15**, 100333. <https://doi.org/10.1016/j.ynstr.2021.100333>
- Zhu X, Fu B, Dong M, Guo Y, Cao Z and Wu J (2021) Effects of long-term antibiotic treatment on mice urinary aromatic amino acid profiles. *Bioscience Reports* **41**(1), BSR20203498. <https://doi.org/10.1042/BSR20203498>

Cite this article: Lampeter T, Love C, Tang TT, Marella AS, Lee HY, Oganyan A, Moffat D, Kareem A, Rusling M, Massmann A, Orr M, Bongiorno C and Yuan L-L 2023. Risk of bias assessment tool for systematic review and meta-analysis of the gut microbiome. *Gut Microbiome*, **4**, e13, 1–20. <https://doi.org/10.1017/gmb.2023.12>