Beyond Probiotics: The Challenge of Changing the Gut Microbiome

Ken Lassesen, M.Sc. *Microbiome Prescription*
26 Aug 2026

Modifying the human gut microbiome is often presented as a straightforward goal to patients: take a probiotic, add more fiber, or change the diet and expect beneficial bacteria to increase. In practice, the microbiome is a complex ecosystem containing hundreds of bacterial species and strains, along with fungi, viruses, archaea, and their metabolic products. Its composition is shaped by diet, medication exposure, host genetics, immune activity, age, geography, disease, and interactions among microbes. As a result, an intervention that changes one person's microbiome may have little effect—or an unintended adverse effect—in another.

Traditional approaches have largely focused on a relatively small group of easily measured cultured bacteria. Many of these bacteria are available as probiotic organisms, especially species and strains of *Lactobacillus* and *Bifidobacterium*. These products may be useful in specific clinical contexts, but they represent only a small fraction of the organisms found in the gut. In many cases, commercial probiotics do not permanently colonize the intestine. Many commercial probiotics are sourced from animals and not humans. These probiotics may be detectable only while they are being consumed, and their effects depend on the existing microbial community and the host environment. Evidence for benefit is often strain-specific, outcome-specific, and inconsistent across populations and different diagnosis.

A major limitation is that, for many gut bacteria, there are few or no well-controlled human studies identifying which dietary components, prebiotics, medications, lifestyle changes, or microbial therapies reliably increase or decrease their abundance. Even when an association is reported —for example, that bacterium is more common in people with a particular diet or health status —this does not establish that deliberately changing that organism will improve health. The organism may be a marker of another underlying condition, may depend on other microbes for survival, or may produce different effects depending on its strain and ecological context.

Microbiome research also faces measurement and interpretation challenges. Sequencing methods can identify bacteria at different levels of resolution, often distinguishing a genus but not the specific strain responsible for a biological effect. Relative-abundance results can be misleading because one organism may appear to decline simply because another increased. Small studies, differing sample-collection methods, short follow-up periods,

and incomplete reporting of diet and medication further limit the ability to compare findings and develop reliable interventions.

Consequently, microbiome modification should not be viewed as a simple process of adding "good bacteria" or eliminating "bad bacteria." More progress will require controlled longitudinal studies, strain-level identification, functional measurements such as metabolite production, and personalized evaluation of baseline microbiome composition, diet, symptoms, and medication exposure. Until that evidence exists, efforts to alter less-studied bacteria remain largely exploratory, and claim that a supplement or dietary intervention can reliably target a specific organism should be treated cautiously.

Below is a suggestion for approaching this challenge until superior data is available.

The Hallucination Model

A clinician receives a microbiome report and sees that 5 bacteria are outside of reference ranges according to the report. The clinician then searches the biomedical literature for an intervention on the US National Institute of Health that appears to move each of those taxa toward the report's reference range and prescribes that intervention. This is prescribed to the patient. Mission accomplished.

Of course, any experienced clinician will know that he will not find such a study. A well-read clinician may be ROLF with this approach for a variety of reasons:

- Reference ranges are typically done using means and standard deviation. This imposes an assumption of a normal distribution on the data. Microbiome data often exhibits a skew of 20-30; a skew over 2 excludes the use of a normal distribution. The term "normal" is often misunderstood; applying a casual conversation meaning instead of the statistical meaning.
    - Laboratory reference intervals should not be confused with clinically meaningful targets. A reference interval is generally a statistical description of a selected comparison population, not evidence that a value outside that interval causes disease or that moving the value inward improves patient outcomes.
- Reference ranges are not causal ranges. The causal attribution has sometime been shown to be a constellation of bacteria that are all in reference ranges but interact with each other.
    - In many microbiome studies, the relevant signal may involve community structure, functional capacity, metabolite production, ecological

interactions, diet, medication exposure, inflammation, or other host factors rather than the abundance of one organism in isolation.

- It is extremely unusual for a clinical microbiome test to be the same as that used in any study. The National Institute of Standards and Technology has identified for over 10 years a severe lack of standardization as undermining microbiome research. Their lead has accurately stated in 2019: "***There are currently 97 different ways to analyze the same raw data, and they will give you 97 different answers*** ". You cannot safely assume that the study results apply to the patient test.
    - Microbiome results depend on numerous choices: specimen collection and storage, DNA extraction, sequencing platform, targeted 16S region or shotgun metagenomic protocol, sequencing depth, taxonomic database, bioinformatic pipeline, normalization method, and reporting scale.
- Finding the targeted bacteria in the same study is extremely unlikely. A common response is to find a study in isolation for each bacterium and then synthesize a combination of substances. This assumes complete independence of each substance and its bacteria influence. This is typically false with many substances helpful for one bacterium and contraindicated to another. To be safe, the clinician will need to read every published study for each substance; a time requirement impractical in a clinical setting.
    - Microbial taxa are embedded in ecological networks, and an intervention that increases one taxon may decrease another, alter total microbial biomass, change relative abundances without changing absolute counts, or affect host physiology through pathways unrelated to the selected taxa. Combining interventions based on isolated taxon-level studies can therefore produce contradictory or uninterpretable effects. The evidentiary burden is not merely to find one paper per organism, but to establish that the combined intervention is safe, clinically relevant, appropriately dosed, and beneficial for patients comparable to the individual being treated.
- The impact of a substance on a bacteria may be inconsistent. Baseline microbiome composition and function, diet, lifestyle, antibiotic exposure, age, comorbid disease, and concomitant medications can all influence both microbial response and clinical effect. The direction of change observed in one population may not generalize to another, and a change in microbial abundance is not necessarily accompanied by an improvement in symptoms or disease outcomes.

A frustrated clinician (or patient) may simply ask some Large Language Model for an answer. The response would often be called ***hearsay*** in a court of law.

Building a better model

Abandon all hope of getting a definitive answer! As an alternative to ignoring the existing data, transform the data into a confidence model. The historic term for the model advocated here is a "fuzzy logic expert system". The method of construction is described below.

Building a Database of Known Interventions

This means extracting data from studies by skilled individuals or well vetted extraction tools. The data typically contains this n-tuple of data. Not all data may be available in the published studies.

- The intervention
- Target population (women with diabetes, mice, etc)
- The bacteria altered, typically NCBI Taxon numbers are effective
- The statistical significance found
- The shift vectors
    - Someone at the 10%ile may increase to the 40%ile
    - Someone at the 80%ile may decrease to the 77%ile
- The population vectors
    - Diagnosis
    - Demographics
- The microbiome pipeline used
- Is this citing other studies or original research
- etc

The studies may be listed on the US National Institute of Health, Europe PMC, Scopus, Web of Science etc.

Ideal versus Reasonable

Assuming the database is well annotated, the key question is: how confident are you that an observed shift will also occur in the intended human population? This confidence is often represented on a scale from -1 to 1, where -1 indicates confidence that the effect will occur in the direction opposite to the target shift. Although it reflects the confidence of an outcome, it is not a true probability.

Some researchers choose to include human studies only. This can make an already sparse evidence base even more limited. A more useful question is: *If flaxseed reduces* Akkermansia *in mice, is it reasonable to hypothesize that flaxseed may also reduce* Akkermansia *in humans?* If the finding comes from ducks or chickens instead, is that hypothesis still reasonable?

The answer determines whether evidence from veterinary science should be excluded. That evidence can be especially valuable because animal studies are often more tightly controlled than human studies.

Well Vetted Extraction Tool Criteria

The extraction and encoding tools being advocated must account for the fact that published-study language is often difficult to interpret reliably. A classic example is a statement such as “normalize some bacteria,” while the specific bacteria are listed elsewhere in the paper.

The acceptable tolerance for incorrect or missing machine-extracted data is therefore critical. The following validation criteria are suggested:

- Evaluate the extraction tool against manually extracted data from more than 200 papers.
- Fewer than 2% of manually extracted data elements should be missing or incorrect in the tool’s output.
- At least 98% of newly identified data elements produced by the tool should be confirmed as correct through subsequent manual review.

A hybrid workflow is recommended: use the tool to create and populate draft data records, then have a human reviewer validate, correct, and formally accept them.

Extending the Database by Inference

When studies on *Limosilactobacillus obtusus* are limited, a clinician may infer that interventions known to increase the broader *Limosilactobacillus* genus will also increase *L. obtusus*. That assumption is not universally valid. For example, Mutaflor (*Escherichia*

coli Nissle 1917) has been reported in multiple studies to reduce the overall relative abundance of *Escherichia*, despite being an *Escherichia* strain itself.

The likely relationship between a taxon and its parent node in the taxonomic hierarchy can instead be estimated empirically. Using a large set of healthy microbiome samples processed through the same analytical pipeline, calculate the correlation coefficient and regression slope between the child taxon and its parent after applying a monotonic increasing transformation to relative-abundance values. The reference dataset should reflect the intended clinical population, include relevant annotations such as age and sex, and ideally be drawn from the same population used to interpret clinical results. At least 1000 samples is recommended for the reference dataset. An example is shown below for L. Reuteri.

NCBI	Name	Rank	R^2	Impact Slope
2742598	Limosilactobacillus	genus	0.48638553363026216	0.6672678076013214
1519	Clostridium tyrobutyricum	species	0.33753336394353384	0.008099960398829079
1515	Acetivibrio thermocellus	species	0.3089745107699126	0.07029052743055722

Other pipelines will produce different numbers (unfortunately).

This approach can expand an intervention-evidence dataset from roughly 90,000 directly observed taxon–intervention relationships to approximately 14 million inferred relationships. This may also identify plausible interventions for taxa that have not yet been evaluated directly in clinical or experimental studies.

Factoring in Confidence

Consider the simplest case: assign a base confidence of 1.0 to an observed increase. For each study, assign a confidence of 1.0 to each individual study when there was an increase, a -1.0 when there was a decrease. In practice, these values may be adjusted by additional data vectors, such as study quality, population characteristics, methodology, sample size, and analytical pipeline.

We therefore need an algorithm for aggregating the results of all studies evaluating a particular intervention's effect on a specific bacterial taxon.

The first question is straightforward: how many non-contradictory studies are needed before we can be reasonably confident in the result? A single study has a substantial risk of being incorrect or non-generalizable because of its study population, experimental design, or processing pipeline.

One possible confidence heuristic is:

- Three concordant studies yield a confidence of 1.0.
- Two concordant studies yield a confidence of 0.84, calculated as $\sqrt{0.7}$
- One concordant study yields a confidence of 0.70, calculated as $\sqrt{0.5}$

When contradictory studies are present, several aggregation heuristics are possible. For example, suppose six studies report an increase while three report a decrease. Possible measures of directional confidence include:

$$\frac{6}{6+3} = 0.67$$

or

$$\frac{\log(6)}{\log(6+3)} \approx 0.82$$

The desired output is a confidence estimate representing the likelihood that the intervention will produce the intended directional shift in the target bacterium.

Moving from Reference Ranges to Causality Ranges

To move beyond reference ranges, we need a large collection of samples processed through the same analytical pipeline. Ideally, each sample is annotated with patient demographics—such as age, sex, and other relevant characteristics—along with two types of clinical information:

- **International Classification of Diseases (ICD).** These provide the most objective clinical outcome vector.
- **Reported symptoms and their severity.** This is a more subjective vector and may be incomplete, since patients can normalize chronic symptoms and fail to report them.

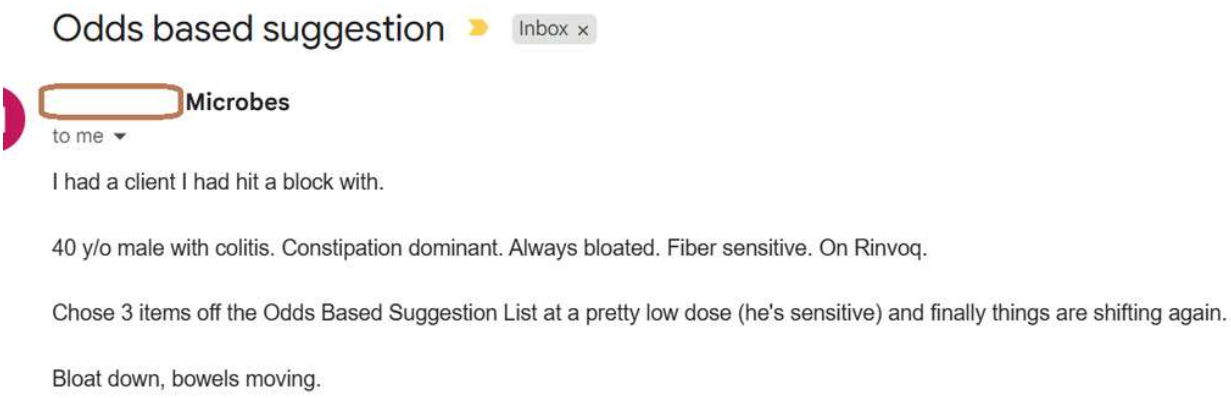
The necessary analyses are more demanding and may require substantial computational time. The preferred method is to calculate **range-based odds ratios**: for each biomarker range, estimate the odds that a participant has a given diagnosis or symptom profile.

A familiar analogy is smoking epidemiology, where the odds of developing lung cancer are evaluated across ranges defined by cigarettes smoked per day and total years of smoking.

A Shortcut for Identifying Causality Ranges

Rather than calculating separate odds ratios for every ICD diagnosis and symptom, we can use the large healthy microbiome reference cohort described above. We calculate the odds that an individual belongs to the unhealthy group versus the healthy reference group. This is symptom independent and avoids the challenge of needing to combine odds ratios from multiple conditions.

This approach has reportedly produced useful results for some clinicians, as illustrated in the email below.


Odds based suggestion Inbox ×

Microbes

to me

I had a client I had hit a block with.

40 y/o male with colitis. Constipation dominant. Always bloated. Fiber sensitive. On Rinvoq.

Chose 3 items off the Odds Based Suggestion List at a pretty low dose (he's sensitive) and finally things are shifting again.

Bloat down, bowels moving.

It is important to note that some high odds ratios may be assigned to bacterial taxa not known to occur in humans. This can happen because the analysis pipeline matches RNA sequences against a broad, generic reference library.

In these cases, the taxonomic label should not necessarily be *interpreted literally*. The meaningful finding is the underlying RNA sequence-pattern match, which may represent an as-yet unidentified or unnamed bacterium, or a related organism that is not well represented in current human microbiome reference databases. The odds-ratio is the pattern, not the bacteria.

Model Summary

From the above we end up with two numbers that need to be combined:

- The Odds Ratio for the bacteria, say 30:1
- The confidence that an intervention will shift in the desired direction

We need to aggregate their products for each targeted bacteria impacted by the intervention. The result is the confidence benefit of the intervention. A simple sum aggregation is a reasonable starting point.

Where should you get this data?

The ideal would be for the microbiome testing company to provide all of the above information to the clinician. The company decides the processing of the sample which cascades into Causality Ranges that are specific for the microbiome test. Similarly, the company should provide the inference table that may be applied to the suggestions.

An example of what a report may look like is here (this is with real data). **Note:** every intervention is linked to sources.