Table 1 Key issues to be addressed in publications applying metagenomics

- Specimen collection, handling, preservation and storage
- Nucleic acid extraction
- Sequencing instrumentation and processing, including library preparation
- Bioinformatic analysis method, including workflow, database composition and parameterisation.
- Quality assurance measures, including internal quality control, such as the use of adequate internal and external controls
- Limits of detection, including analytical sensitivity and specificity for clinical testing
- Power and sample size calculations
- Use of orthogonal methods to confirm sequencing results
- Criteria to confirm the role of pathogen(s) in disease aetiology
- Turnaround time
- Cost
- Ethical considerations
- Specific issues related to applications, such as in the diagnosis of CNS infections, and investigation of AMR

Table 2. Examples of potential sources of bias in metagenomics studies and implications for result interpretation. This list is not comprehensive, but illustrates how results may be affected by collection, processing and analysis methods.

Potential source of bias	Example of implication for results interpretation
Specimen collection methods	Collection without a cold chain or nucleic acid stabilising agents may cause nucleic acid degradation and potential false negative results or overgrowth of selected organisms leading to misinterpretation of abundance; multiple freeze-thawing may also cause nucleic acid degradation.
Nucleic acid extraction method	A lack of bead-beating step may limit the detection of difficult-to-lyse bacteria; small specimen volumes may reduce the ability to detect low-level organisms.
Sequencing library preparation	Poly-A tail enrichment of RNA will not include fragmented pathogen genomes; DNA sequencing alone will not detect RNA viruses.
Targeting of sequences	Capture probe-targeted sequencing will limit detection to targeted, known sequences. 16S targeted sequencing as opposed to whole genome sequencing will have limitations with regard to the level of taxonomic classification.
Sequencing methods	High-level sample multiplexing may lead to insufficient read depth to allow detection of organisms present at low levels. Computational contamination may occur between samples pooled on the same sequencing run due to a sample barcode for a sequence being misread and misassigned to another sample on the same run (82). This is termed ‘barcode bleed-through’, and dual barcodes drop the rate of bleed through dramatically compared to single barcodes. Unique molecular identifiers (UMIs) are an even more powerful way to identify this phenomenon.
Processing controls	Negative controls allow some contaminating organisms to be identified. Internal positive controls, reference standards such as sequins, reduce bias introduced by experimental variability, and may improve recognition of low-level organisms.
Analysis methods	A small curated database, or highly stringent criteria may limit the identification of novel or unexpected organisms, leading to false negative results; an un-curated database or lenient criteria may mis-identify organisms.