Critical Care

REVIEW

Open Access

Consistency between metagenomic next-generation sequencing versus traditional microbiological tests for infective disease: systemic review and meta-analysis

Chengxi Liu[1†], Xiao Song[1†], Jihai Liu[1], Liang Zong[1], Tao Xu[2], Xu Han[1], Fan Li[1], Bo Li[1], Huadong Zhu[1*] and Di Shi[1*]

Abstract

Background Pathogen identification is essential in sepsis and septic shock. Metagenomic next-generation sequencing (mNGS) is a novel pathogen detection method with several advantages over traditional tests. However, the consistency between mNGS and traditional pathogen tests requires further investigation.

Objectives We aimed to assess the consistency between mNGS and traditional pathogen tests and to identify the factors influencing this consistency.

Methods This systematic review and meta-analysis involved a comprehensive search of mNGS and traditional pathogen tests in PubMed, Embase, Scopus, Web of Science, and the Cochrane Library. Data from included studies were extracted, and kappa consistency between mNGS and traditional tests was calculated. Study quality was evaluated using the QUADAS-2 tool.

Results The search identified 415 studies, of which 27 were included in the analysis, involving 4112 individuals. Meta-analysis showed a pooled consistency of 0.319 ± 0.013 ($p < 0.001$), indicating a moderate relationship. In terms of sample type, cerebrospinal fluid showed the highest pooled kappa consistency at 0.500 ± 0.029 ($p < 0.001$). Immunocompromised patients had a lower pooled kappa consistency of 0.294 ± 0.014 ($p < 0.001$) compared to 0.321 ± 0.028 ($p < 0.001$) in immunocompetent patients. Positive percent agreement of mNGS was 83.63% over traditional microbiological test, and negative percent agreement was 54.59%.

Conclusion This review demonstrates a moderate relationship between mNGS and traditional pathogen tests, indicating a complex relationship between these two methods. Sterile samples show higher consistency than non-sterile samples. Immune function deficiency may reduce the consistency between mNGS and traditional tests. Further research is needed on the use of mNGS in sepsis and septic shock.

Keywords mNGS, Immunocompromised, Kappa consistency

†Chengxi Liu and Xiao Song have contributed equally to this work.

*Correspondence:
Huadong Zhu
zhuhuadong1970@126.com
Di Shi
sddidizy@126.com
Full list of author information is available at the end of the article

Introduction

Sepsis, caused by severe infections from various pathogenic microorganisms, is a leading cause of hospitalization and mortality worldwide [1, 2]. Approximately 1.7 million cases of sepsis occur annually, resulting in nearly 250,000 deaths, making it the leading cause of death in noncardiac ICUs [3, 4]. The inappropriate use of antibiotics in the early stages is a major contributor to the increased morbidity and mortality associated with sepsis [5]. Previous studies suggest that each hour of delayed antibiotic administration raises the absolute mortality rate by about 0.3 to 1.8% [6].

Early and accurate pathogen identification helps clinicians choose targeted treatments, improving patient outcomes and reducing mortality [7, 8]. However, there are no clear guidelines regarding the optimal method for pathogen identification. Traditional microbiological tests, such as smear, culture, serology, and limited molecular panels, fail to detect the full range of potential pathogens, including fastidious, nonculturable, or rare organisms not covered by current panels. Additionally, culture methods in many clinical laboratories are time-consuming, typically requiring 3–5 days for results. Serological tests and PCR require prior assumptions about the presence of specific pathogens and the availability of the necessary commercial kits, leading to missed detections in some cases [2].

Metagenomic next-generation sequencing (mNGS) is a novel technology for pathogen detection in infectious diseases and sepsis. It combines high-throughput sequencing with bioinformatics analysis [9]. mNGS can simultaneously detect multiple pathogens through DNA or RNA sequencing of clinical samples [10]. Studies have shown that mNGS has higher sensitivity and can identify a broader range of pathogens compared to traditional microbiological tests [11, 12]. It has also demonstrated superior performance in detecting mixed pulmonary infections compared to conventional methods [13].

Although numerous studies have focused on the sensitivity and specificity of mNGS [14–18], few have investigated the correlation between mNGS and traditional testing methods. Our literature review of studies comparing mNGS and traditional microbiological methods was conducted. Our literature review attempts to establish a connection between traditional pathogen detection and mNGS through the Cohen's kappa value, making it easier for clinicians to understand the results of mNGS. Cohen's Kappa was developed by a prominent statistician called Jacob Cohen in 1960s [19]. Cohen's kappa was developed as a tool to evaluate the extent of agreement among different raters, which also called inter-rater reliability. In this article, we used Cohen's Kappa to assess the consistency between mNGS and traditional microbiological tests in detecting pathogens in infected patients. Higher consistency between these two methods indicates a higher level of reliability for both methods in pathogen detection. Any factors that may reduce consistency will affect the reliability of both methods in microbiological detection. At the same time, high consistency also suggests that the two methods can be interchangeable in specific situations. This review aims to assess the consistency between mNGS and traditional microbiological methods and to compare consistency across different study types, sample types, and immune statuses.

Methods

This systematic review and meta-analysis followed the PRISMA2020 guidelines (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and was pre-registered on PROSPERO with registration number CRD42024507424.

Inclusion criteria

The study included all research comparing traditional microbiological testing and metagenomic next-generation sequencing (mNGS) across all available sample types (in our article includes bronchoalveolar lavage fluid, joint fluid, urine, cerebrospinal fluid, and plasma) in infectious disease. Traditional microbiological testing include smear, staining, culture, serology test, and limited molecular panels. Studies such as randomized controlled trials, retrospective cohorts, prospective cohorts, case–control studies and observational studies were eligible. The patient population consisted of individuals aged 18 years or older with infectious diseases. The infectious disease diagnosis defined according to the clinical comprehensive diagnosis, which was made by different independent clinicians with all clinical data [20].

Exclusion criteria

Articles were excluded if they met the following criteria: (1) reviews, letters, case reports, or case series involving fewer than 10 patients; (2) studies involving non-human subjects; (3) studies where enrolled patients did not have infectious diseases; (4) studies that did not compare results between mNGS and traditional methods; (5) low-quality studies or those without relevant data.

Literature search

A search was conducted in MEDLINE via PubMed, Embase, Web of Science, and the Cochrane Library for English and Chinese language articles published up to September 2024. The search terms included "Metagenomic Next-Generation Sequencing", "mNGS", "infections" and "trial". Both retrospective and prospective studies comparing traditional microbiological testing with

mNGS were included. In addition to electronic databases, we also explored other sources of literature, including conference proceedings, grey literature, and clinical trial registries to minimize publication bias.

Selection of studies

Four independent reviewers (XS, LZ, CXL, and HJ) evaluated the titles and abstracts of identified studies. Full texts were then assessed for eligibility. Potentially relevant studies cited in the reference lists of identified articles and the studies cited the reference articles were also reviewed. Any conflicts between reviewers were resolved by an additional experienced reviewer (DS).

Data extraction

Data were independently extracted by four reviewers (XS, LZ, CXL, and HJ). From each study, the following data were collected and summarized: (1) basic characteristics including authors, study period, and year of publication. (2) Type of study design, sample type, infectious diagnosis at the time of testing, and immune status of the patients. immunocompromised status was defined as: primary immunodeficiency, solid organ transplantation (SOT), haematopoietic stem cell transplantation (HSCT) during the last 6 months, HIV infection with CD4 counts < 200 mm^3, active tuberculosis, hematologic malignancy, neutropenia, cytotoxic chemotherapy during the last month, autoimmune disease requiring glucocorticoid treatment or immunosuppressive therapy for more than 2 weeks. (3) Original data from mNGS and traditional microbiological tests for calculating Kappa consistency, which is the positive or negative results of each experimental individual in mNGS and traditional microbiological tests.

Outcomes

The primary outcome was the diagnostic consistency between mNGS and traditional microbiological tests, evaluated using Kappa consistency (also called Cohen's Kappa). Kappa consistency was calculated using the following formula:

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

The formula for kappa consistency is defined as follows, where p_o is the relative observed agreement between raters, and p_e is the hypothetical probability of chance agreement. pe is calculated based on the probabilities of each observer randomly assigning each category, using the observed data. According to previous studies [21], kappa values are interpreted as indicating no reliability (< 0.01), slight reliability (0.01–0.2), fair reliability (0.21–0.40), moderate reliability (0.41–0.60), substantial reliability (0.61–0.80), or almost perfect reliability (0.81–1.00).

The secondary outcomes included the positive percent agreement (PPA) and negative percent agreement (NPA), with mNGS treated as the test method and traditional microbiological tests treated as the reference method.

Quality assessment

The quality of the included studies was assessed by four independent reviewers (XS, LZ, CXL, and HJ) using the QUADAS-2 tool (Quality Assessment of Diagnostic Accuracy Studies) in RevMan5 software (via the QUADS2 checklist). We treated mNGS as index test and traditional microbiological methods as reference standard. Disagreements were resolved through consensus meetings with an experienced reviewer (DS).

Statistical analysis and data synthesis

The random-effects was used to pool the proportions of consistency and SE(P) between mNGS and traditional microbiological tests. Heterogeneity among studies was evaluated using the χ^2 test (Q test) and the I^2 statistic.

All p-values were based on two-sided tests, with p-values < 0.05 considered statistically significant. Data analyses were performed using R software version 4.1.0 (http://www.R-project.org) and Review Manager (RevMan, version 5.3, The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, 2014).

Results

Selection of studies

The details of the search strategy and study selection process are presented in Fig. 1. We created this flow diagram according to PRISMA. (Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimized digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230) A total of 415 studies were identified through the search. After screening, retrieval, and eligibility assessment, 27 studies reporting original data from both mNGS and traditional tests were included in this review.

Study characteristics

The characteristics of the 27 included studies are summarized in Table 1, and included studies is listed in Supplementary File 1. All were observational, with 17 designed retrospectively and 10 designed prospectively. Notably, none of the studies are cross-sectional design nor case–control design. Geographically, most of the studies were conducted in China, highlighting a significant focus on this region. Only one study was conducted in France,

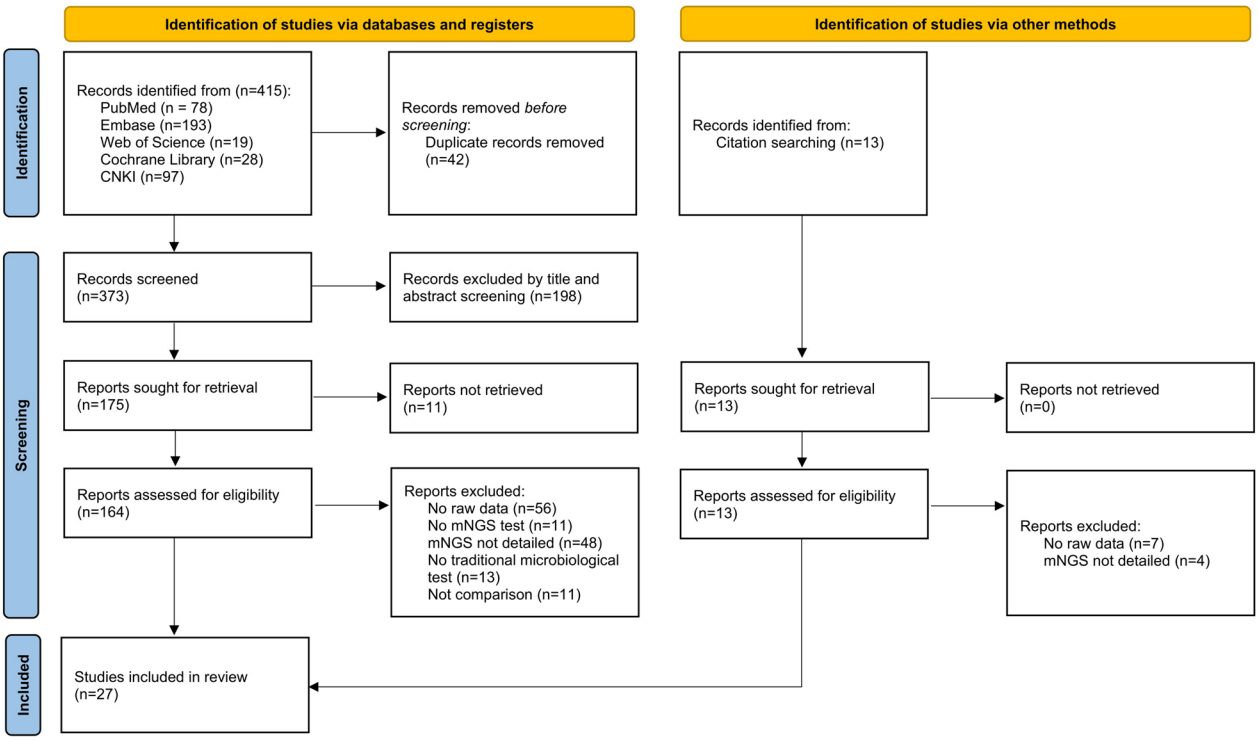

Fig. 1 Research flow

and another in the Czech Republic, reflecting the predominantly Chinese-based research in this area. In terms of scope, six studies involved multiple centers, while 21 studies were conducted in single center, focusing on localized populations. The total sample size is 4112 participants, and average sample size across the studies was 152.3 participants. The average age of participants was 54 (13–93) years, with only one study not reporting the age of its population. Regarding the technologies used for genetic sequencing, seven studies used the BGISEQ platform, fourteen used the Illumina platform, two used MinION platform, two used MGISEQ platform, one used Dinfectome Dif seq platform, one utilized untargeted next-generation sequencing, and one used 16S rRNA PCR-DGGE-S. One uniquely employed both the BGISEQ and Illumina platforms simultaneously. This diversity in approaches underscores the complexity and range of methodologies used in the research, reflecting a comprehensive examination of the subject matter across different technologies and platforms.

Consistency and subgroup analysis

A total of 27studies, involving 4112 individuals, were included in this analysis. The pooled kappa consistency was calculated to be 0.319 ± 0.013, with a highly significant p-value of less than 0.001, as shown in Fig. 2. The variability among the studies was evaluated, resulting in an I^2 statistic of 90.15%, indicating a substantial degree of heterogeneity in the data. Additionally, the tau^2 value was recorded as 0.07, which helps to explain the variation across the included studies. Further details are provided in the subgroup analysis in Table 2.

In terms of study design, the pooled kappa consistency for prospective studies was 0.181 ± 0.015 ($p < 0.001$). In contrast, retrospective studies showed a lower pooled kappa consistency of 0.423 ± 0.021 ($p < 0.001$).

Of the 27 studies analyzed, 15 focused on bronchoalveolar lavage fluid (BALF), 1 focused on sputum. Three studies examined cerebrospinal fluid (CSF), while 3 studies used plasma samples. Two study focused on joint fluid, one study focused on biopsy tissue, one study focused on ascites, and one collected urine samples. The pooled kappa consistency for BALF was found to be 0.225 ± 0.019 ($p < 0.001$), and the pooled kappa consistency for sputum was 0.045 ± 0.012 ($p = 0.005$). For joint fluid, the pooled kappa consistency was 0.088 ± 0.070 ($p = 0.184$), suggesting some consistency, although not statistically significant. In the case of urine samples, the pooled kappa consistency was 0.438 ± 0.167 ($p = 0.01$), indicating significant findings. Studies involving cerebrospinal fluid demonstrated a pooled kappa consistency of 0.500 ± 0.029 ($p < 0.001$), reflecting strong consistency. However, plasma samples showed a lower pooled kappa consistency of 0.161 ± 0.040 ($p = < 0.001$). The pooled

Table 1 Summary of included studies

Study name	Research type	Centers	Region	Sample size	Male (%)	Age	Sample type	Immuno-compromised	Diagnosis	mNGS method	traditional microbiological tests
Parize 2017	Prospective	Multiple	France	101	48.51	57 (19–93)	Plasma	Yes	Bloodstream infection	Untargeted next-generation sequencing	Culture and serological test
Kotaskova 2019	Prospective	Single	Czech	29	74.44	69 (32–89)	Urine	No	Acute cystitis	16S rRNA PCR-DGGE-S	Culture
Zhang 2020	Prospective	Single	China	142	60.43	43 (13–84)	CNF	No	CNS infections	BGISEQ-100 platform	Smear, culture and serological test
Xing 2020	Prospective	Multiple	China	1023	36.15	41 (14–80)	CNF	No	CNS infections	BGISEQ-500/50 platform	Smear and culture
Huang 2020	Prospective	Single	China	104	58.65	–	Joint Fluid	No	Prosthetic joint infection	BGISEQ-500 platform	Culture
Chen 2022	Prospective	Single	China	156	49.48	41 (13–80)	CNF	No	CNS infections	BGISEQ-50 platform	Smear, culture and PCR
He 2022	Retrospective	Multiple	China	151	64.24	55 (14–83)	BALF	No	Pulmonary infection	Illumina NextSeq CN500 platform	Culture and PCR
Fu 2022	Retrospective	Multiple	China	54	65.96	38 (18–76)	Plasma	Yes	Bloodstream infection	Illumina NextSeq 550Dx platform	Culture
Ju 2022	Retrospective	Single	China	163	84.11	56 (25–82)	BALF	Yes	Pulmonary infection	Illumina NextSeq 550Dx platform	Culture
Lu 2022	Retrospective	Single	China	134	64.93	65 (31–80)	BALF	No	Pulmonary infection	Illumina NextSeq CN500 platform	Culture and serological test
Dong 2023	Retrospective	Single	China	106	43.40	56 (15–78)	BALF	No	Pulmonary infection	Illumina SEQ sequencing platform	Culture
Pan 2023	Retrospective	Single	China	79	49.02	58 (22–80)	BALF	No	Pulmonary infection	BGISEQ-200 and Illumina NextSeq550 platform	Smear and culture
Chen 2023	retrospective	Single	China	64	71.43	70 (35–87)	BALF	No	Pulmonary infection	MinION platform	Culture
Liu 2023	Retrospective	Single	China	162	61.11	55 (20–82)	Plasma	No	Bloodstream infection	Illumina NextSeq 550 platform	Culture
Xie 2023	Retrospective	Single	China	57	91.23	44 (39–76)	BALF	Yes	Pulmonary infection	Illumina NextSeq 550 platform	Smear and culture
Zhang 2023	Retrospective	Single	China	184	64.67	62 (30–89)	BALF	Yes	Pulmonary infection	BGISEQ-50 platform	Culture
Zhao 2023	Retrospective	Single	China	57	20.69	43 (14–73)	BALF	Yes	Pulmonary infection	Illumina NextSeq 550Dx platform	Smear, staining and culture
Wen 2023	Retrospective	Single	China	40	37.50	51 (19–78)	BALF	Yes	Pulmonary infection	BGISEQ-500 platform	Smear and culture
Lin 2024	Prospective	Multiple	China	168	60.09	56 (16–90)	BALF	No	Pulmonary infection	Illumina Nextseq550 platform	Smear, staining and culture
Fu 2024	Retrospective	Multiple	China	140	68.57	34 (15–76)	BALF	Yes	Pulmonary infection	Illumina Nextseq platform	Smear and culture
Zhang 2024	Prospective	Single	China	41	48.00	51 (32–72)	Ascites	No	Peritonitis	Illumina NextSeq 550Dx platform	Culture
Jiang 2024	Prospective	Single	China	56	42.72	59 (18–70)	BALF	No	Pulmonary infection	MGISEQ-200 platform	culture

Table 1 (continued)

Study name	Research type	Centers	Region	Sample size	Male (%)	Age	Sample type	Immuno-compromised	Diagnosis	mNGS method	traditional microbiological tests
Zheng 2024	Retrospective	Single	China	155	42.47	58 (29–82)	BALF	No	Pulmonary infection	MGISEQ-2000 platform	Culture
Lv 2024	Retrospective	Single	China	59	61.84	59 (29–82)	Biopsy tissue	No	Spinal infection	Dinfectome Dif seq platform	culture
Liu 2024	Retrospective	Single	China	516	48.83	58 (43–77)	Sputum	No	Pulmonary infection	MinION platform	Culture
Chen 2024	Prospective	Single	China	125	61.60	55 (21–91)	BALF	No	Pulmonary infection	Illumina NextSeq platform	culture
Shi 2024	Retrospective	Single	China	46	47.83	68 (28–89)	Joint Fluid	No	Prosthetic joint infection	Illumina NextSeq platform	Culture
Total				4112	56.59	54 (13–93)					

CNF = cerebrospinal fluid, CNS = central nervous system, BALF = bronchoalveolar lavage fluid, PCR = polymerase chain reaction

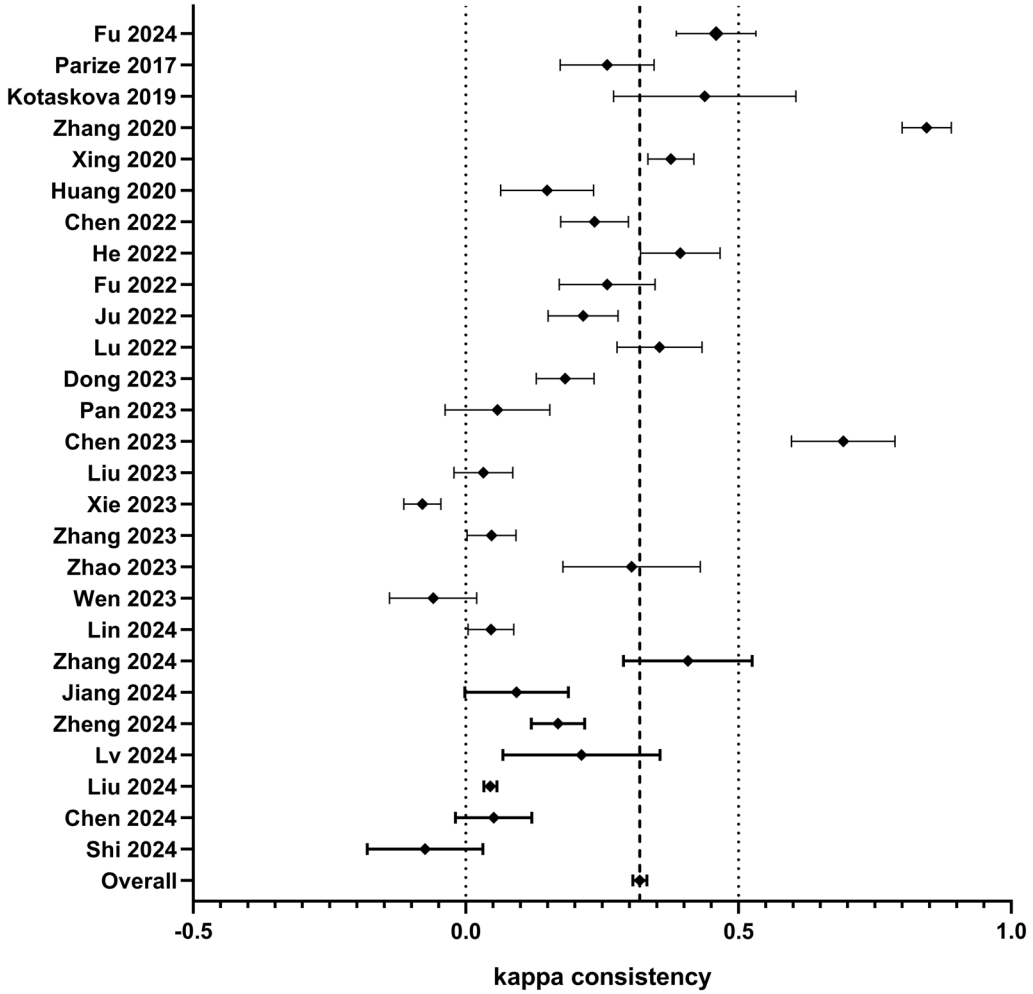

Fig. 2 Pooled kappa consistency

kappa consistency of biopsy tissue was 0.212 ± 0.144 ($p = 0.102$). And the pooled kappa consistency of ascites was 0.407 ± 0.118 ($p = 0.003$), indicate also a significant finding.

When consider the infectious diagnosis of infectious disease, it's similar to that of different sample type. Sixteen studies focused on pulmonary infection, of which pooled kappa consistency is 0.165 ± 0.015 ($p < 0.001$). Three studies focused on CNS (central nervous system) infection, of which kappa consistency is 0.500 ± 0.029 ($p < 0.001$). Three studies focused on bloodstream infection, of which kappa consistency is 0.161 ± 0.040 ($p < 0.001$). Two studies reported prosthetic joint infection with kappa consistency is 0.088 ± 0.070 ($p = 0.184$). The kappa consistency of studies concerned peritonitis, spinal infection and acute cystitis are separately the same as it of studies concerned ascites, biopsy tissue and urine.

Finally, when analyzing the immune status of patients, the pooled kappa consistency of eight studies focusing on immunocompromised patients was 0.294 ± 0.014 ($p < 0.001$). In comparison, nineteen studies involving healthy patients had a similar pooled kappa consistency of 0.321 ± 0.028 ($p < 0.001$). This detailed breakdown provides valuable insights into the variations observed across different patient groups and sample types within the studies included in this analysis.

Secondary outcome

Our study also calculated PPA and NPA with mNGS treated as index test and traditional microbiological tests treated as reference method, which are presented in Table 3. When using the traditional method as the index test, the pooled positive percent agreement was 83.63%, the pooled negative percent agreement was 54.59%, the pooled false positive rate was 45.41%, and the pooled false negative rate was 16.37%.

Table 2 Subgroup analysis

Subgroup factor	Number of studies	Kappa consistency (Mean ± SD)	*p*-value
Total	27	0.319 ± 0.013	< 0.001
Type of research			
Prospective	10	0.181 ± 0.015	< 0.001
Retrospective	17	0.423 ± 0.021	< 0.001
Sample Type			
BALF	15	0.225 ± 0.019	< 0.001
Plasma	3	0.161 ± 0.040	< 0.001
CNF	3	0.500 ± 0.029	< 0.001
Joint Fluid	2	0.088 ± 0.070	0.184
Sputum	1	0.045 ± 0.012	0.005
Biopsy tissue	1	0.212 ± 0.144	0.102
Ascites	1	0.407 ± 0.118	0.003
Urine	1	0.438 ± 0.167	0.01
Immunocompromised			
Yes	8	0.294 ± 0.014	< 0.001
No	19	0.321 ± 0.028	< 0.001
Infectious diagnosis			
Pulmonary infection	16	0.165 ± 0.015	< 0.001
CNS infection	3	0.500 ± 0.029	< 0.001
Bloodstream infection	3	0.161 ± 0.040	< 0.001
Prosthetic joint infection	2	0.088 ± 0.070	0.184
Peritonitis	1	0.407 ± 0.118	0.003
Spinal infection	1	0.212 ± 0.144	0.102
Acute cystitis	1	0.438 ± 0.167	0.01

BALF = bronchoalveolar lavage fluid, CNF = cerebrospinal fluid, CNS = central nervous system,

Table 3 Agreement of the included studies

Study name	PPA	NPA	FP	FN
Parize 2017	81.82	70.00	30.00	18.18
Kotaskova 2019	73.91	83.33	16.67	26.09
Zhang 2020	93.85	90.91	9.09	6.15
Xing 2020	45.80	91.93	8.07	54.20
Huang 2020	90.63	22.50	77.50	9.38
Chen 2022	83.64	44.55	55.45	16.36
He 2022	81.08	58.44	41.56	18.92
Fu 2022	100.00	57.45	42.55	0.00
Ju 2022	92.31	27.78	72.22	7.69
Lu 2022	86.67	47.46	52.54	13.33
Dong 2023	94.12	44.94	55.06	5.88
Pan 2023	60.87	46.43	53.57	39.13
Chen 2023	90.24	78.26	21.74	9.76
Liu 2023	72.09	32.77	67.23	27.91
Xie 2023	94.00	0.00	100.00	6.00
Zhang 2023	94.23	10.00	90.00	5.77
Zhao 2023	91.89	35.00	65.00	8.11
Wen 2023	88.89	4.55	95.45	11.11
Lin 2024	87.93	18.18	81.82	12.07
Fu 2024	91.03	53.23	46.77	8.97
Zhang 2024	92.86	55.56	44.44	7.14
Jiang 2024	88.89	20.69	79.31	11.11
Zheng 2024	93.44	26.60	73.40	6.56
Lv 2024	80.85	41.67	58.33	19.15
Liu 2024	93.22	22.76	77.24	6.78
Chen 2024	70.00	36.47	63.53	30.00
Shi 2024	87.10	6.67	93.33	12.90
Total	83.63	54.59	45.41	16.37

PPA = positive percent agreement, NPA = negative percent agreement, FP = false positive, FN = false negative

Quality assessment

Most of the studies included in this analysis had a low risk of bias in the reference standard, patient selection, and flow and timing (Fig. 3). Only one study was rated as having an unclear risk due to a lack of clear patient selection criteria, testing method, and timing details. However, since this study provided clear data on mNGS and traditional pathogen testing, it was included in the analysis.

The funnel plot (Fig. 4) was used to assess publication bias, and the linear regression test for asymmetry showed low publication bias among the included studies ($t = -1.365$, $df = 17$, $p = 0.190$).

Discussion

In this systematic review, we compiled various studies reporting the use of mNGS and traditional testing methods for pathogen identification in infectious diseases. Our study carefully extracted the original data from these studies and calculated the pooled kappa consistency between mNGS and traditional pathogen testing methods, ensuring accuracy and reliability.

Previous studies have found that mNGS has significant advantages over traditional pathogen detection in terms of speed, breadth, and accuracy [20, 22–24]. This research aimed to describe the agreement between mNGS and traditional microbiological teat. The measurement we used was Cohen kappa consistency. Kappa is typically used to describe the reliability between different raters measuring the same variable. This interrater reliability is also called agreement [25]. The agreement may not determine which method is better; rather, it is used to describe the degree of similarity between the two methods.

The pooled kappa consistency demonstrated a moderate level of reliability between mNGS and traditional pathogen tests. This indicate there is a positive correlation between mNGS and traditional pathogen tests, while the degree of this correlation is medial. Notably,

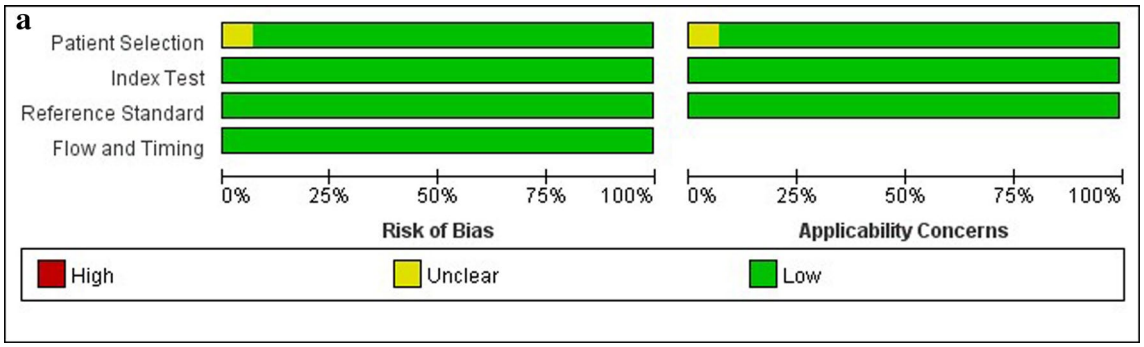

Fig. 3 **a** Quality assessment summary. **b** Quality assessment detail

no comprehensive meta-analysis has specifically focused on the kappa consistency between mNGS and traditional pathogen testing; however, a limited number of independent studies have explored this topic, as referenced in sources [26–31]. Furthermore, the kappa consistency varied depending on the sample type and patient categories included in the studies. This variability highlights the complexity of interpreting results across different methodologies and patient populations and underscores the need for further research to fully understand the relationship between these diagnostic approaches.

Regarding sample types, the kappa consistency for samples from the respiratory tract, cerebrospinal fluid (CSF), and plasma showed statistical significance, though the level of heterogeneity across the studies was high. Among the sample types, CSF exhibited the highest kappa consistency, which aligns with findings from previous studies [30]. This higher consistency may be due to the fact that CSF is typically sterile under normal physiological conditions, making pathogen detection more reliable. In contrast, the kappa consistency for respiratory samples has been shown in earlier studies [26, 29, 31] to range from low to moderate. This lower consistency could be attributed to the large number of commensal microorganisms naturally present in the respiratory tract, suggesting that while mNGS holds promise, it may not yet be fully capable of replacing traditional methods for diagnosing respiratory infections. The significant heterogeneity observed across different sample types emphasizes the need for further research to optimize metagenomic sequencing techniques for various clinical scenarios. This optimization could improve diagnostic accuracy and reliability, ultimately enhancing patient outcomes in diverse healthcare settings, especially in sepsis and septic shock. Current guidelines for septic shock recommend obtaining non-invasive microbiological specimens, such as blood, cerebrospinal fluid, urine, wounds, respiratory secretions, and other body fluids, before initiating antimicrobial therapy [31].

According to our study, the consistency of microbiological specimens from non-sterile sites, such as the respiratory and gastrointestinal tracts, is relatively low. This may suggest that obtaining physiologically sterile specimens, like plasma and cerebrospinal fluid, could provide more accurate pathogen information compared to non-sterile site samples. For patients with sepsis or septic shock, this may imply that blood cultures are more valuable than sputum or urine cultures. Furthermore, for the high kappa consistency in CNF sample, the mNGS and traditional culture may lead to same result, which could save the cost for clinical practice. Previous studies have also shown that using multiple sample types can increase the detection rate and accuracy of pathogen identification [32, 33]. The difference between traditional tests and mNGS in different sample types could affect diagnostic accuracy. Although traditional culture remains the gold standard for pathogen identification, mNGS results may offer valuable insights, particularly in complex infections like sepsis or septic shock [34].

Additionally, this meta-analysis revealed significantly lower kappa consistency for studies focusing on immunocompromised patients compared to those involving immunocompetent individuals. This finding has not been reported in previous studies and highlights a gap in the literature. A possible explanation is the increased diversity and prevalence of opportunistic pathogens in immunodeficient populations, which complicates diagnosis. In immunocompromised patients, the immune system is less effective in fighting infections, leading to a wider range of potential pathogens that need to be identified. Traditional culture-based methods often struggle to detect all infectious organisms in these patients due to the limitations of these techniques in capturing a diverse microbial landscape. In contrast, mNGS offers

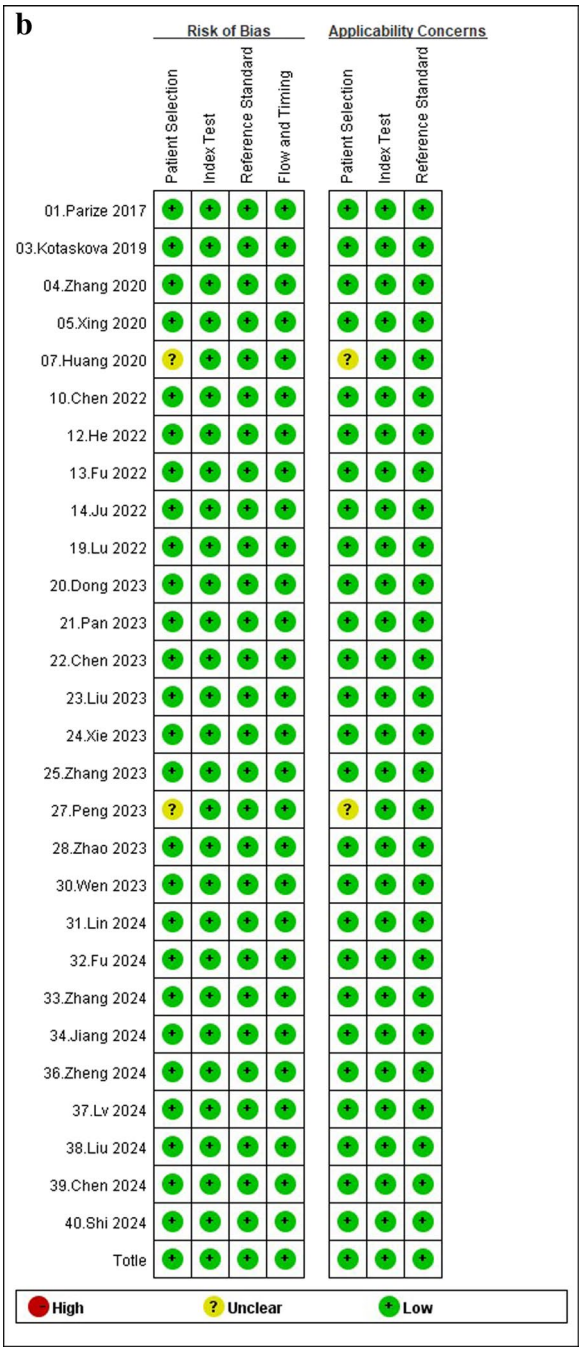

Fig. 3 continued

a significant advantage by detecting a broader range of microorganisms, making it particularly useful for identifying pathogens in complex cases associated with immunosuppression. This highlights the potential of mNGS as a more comprehensive diagnostic tool, especially in situations where traditional methods may be inadequate, thus improving patient management and treatment outcomes in vulnerable populations. According to previous studies, sepsis and septic shock can suppress the immune competence of patients. Traditional cultures may be affected by the diversity of microorganisms due to the immunosuppressive state during sepsis. However, mNGS has a much broader pathogen spectrum [11, 12], making it potentially more useful in diagnosing infections

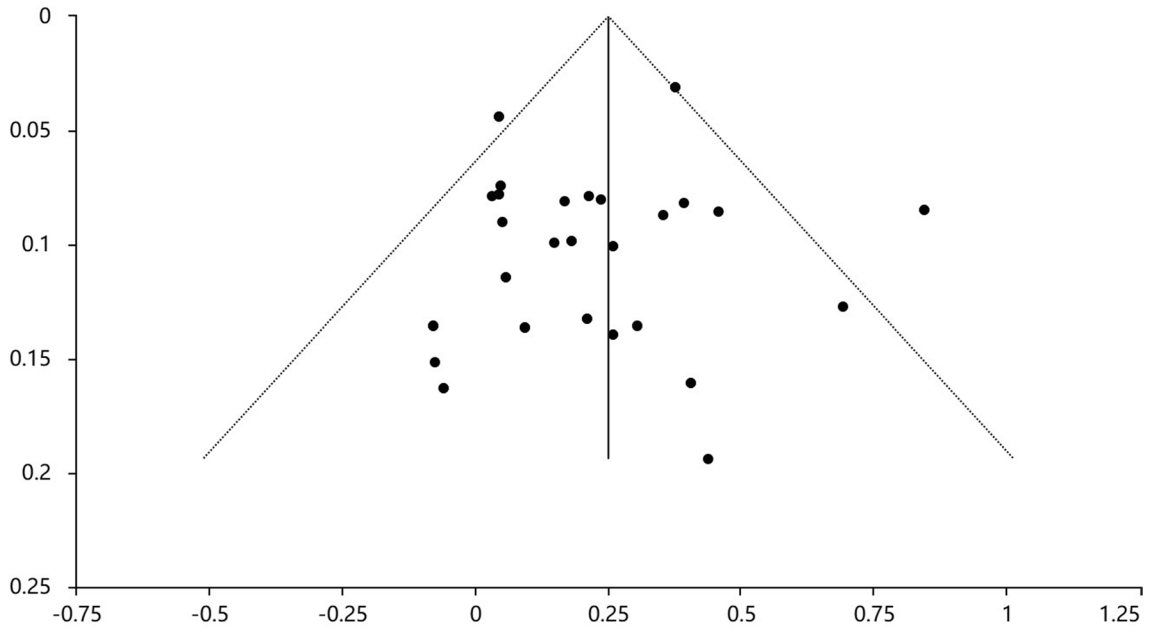

Fig. 4 The funnel plot

in patients with sepsis or septic shock. For uncommon pathogens, previous studies have demonstrated that mNGS has higher diagnostic accuracy than traditional tests for tuberculosis, fungi, and certain viruses [35–37]. This suggests that mNGS may have greater potential for detecting uncommon pathogens compared to traditional tests. Additionally, because mNGS results are available more quickly than traditional cultures, it could reduce the time between diagnosis and antibiotic administration, which could improve survival rates in patients with septic shock.

To our knowledge, no comprehensive meta-analysis has been conducted on the kappa consistency between mNGS and traditional testing methods for infectious diseases. This review represents an important advancement in evaluating kappa consistency, providing valuable insights into the strength of the relationship between mNGS and traditional tests. As we described above, Cohen's Kappa could evaluate interrater reliability. By quantifying this relationship, the findings offer a clearer numerical understanding of how these two diagnostic approaches correlate, especially in cases where clinical results may appear contradictory. Such discrepancies can create confusion in patient management and treatment decisions, making it essential to have a clearer framework for interpreting the consistency between these methods. This analysis not only highlights the strengths and limitations of each approach but also serves as a crucial resource for clinicians facing complex diagnostic scenarios. By shedding light on the nuances of kappa consistency, this review aims to improve decision-making processes in clinical settings, thereby enhancing patient care and outcomes in sepsis and septic shock.

In our research, the mNGS and traditional microbiological test has a medial kappa consistency. This indicate that although these two pathogen detection methods have positive correlation, they could not replace each other in the clinical diagnosis of infectious disease. As for the PPA and NPA between these two methods, we found that PPA is much higher than NPA. This suggest that these two tests have better reliability when the test result are positive. It reminds us that in clinical practice, if one of the tests (mNGS or traditional pathogen detection) is positive, the other test is likely to also be positive. However, if one test returns a negative result, the other test is not necessarily negative. In order to increase the reliability, repeating tests using an alternative method or the same method is recommended. This correlation may help us to decide whether to perform another test when we had the results of one test.

However, there are limitations to our research. First, the included studies were not prospective randomized clinical trials; many were retrospective. Second, we only included studies published in English and Chinese, which may introduce regional bias. Third, some studies were excluded due to unavailable original data, which could also contribute to bias.

Conclusion

In this systematic review, we examined the agreement between mNGS and traditional pathogen testing methods with Cohen kappa consistency. Our analysis showed a moderate kappa consistency, indicating a positive correlation between mNGS and traditional microbiological test, which influenced by factors such as sample type and patient population. Notably, cerebrospinal fluid samples exhibited the highest consistency, likely due to their typically sterile nature, while respiratory tract samples displayed lower reliability due to the presence of numerous commensal organisms. That suggest that mNGS and traditional microbiological tests are alternative when it meet specific infection like CNS infection. Additionally, we found lower kappa consistency in immunocompromised patients compared to immunocompetent individuals, this may be caused by the advantage of mNGS in detecting a broader range of pathogens in complex cases, such as sepsis and septic shock. The medial agreement and high PPA between mNGS and traditional microbiological test indicate that when one test result is positive, the other test is not necessary. However, when one test result is negative, it is recommended to complete the other test to avoid missed diagnoses. Better understanding of the consistency between the two methods and the influencing factors could help choose and interpret results based on clinical needs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13054-025-05288-9.

Additional file1 (DOCX 21 KB)

Additional file2 (DOCX 16 KB)

Author contributions

Chengxi Liu, Xiao Song wrote the main manuscript text and contributed to study design, data collection and analysis. Jihai Liu, Liang Zong, Hui Jiang contributed to data collection and prepared figures. Xu Han, Bo Li, Fan Li worked on table design. Shicheng Ma contributed to text editing and revision of the article. Tao Xu contributed to statistic analysis. Di Shi, Huadong Zhu contributed to study design, study supervision, main manuscript text editing.

Funding

All the authors in this research have no funding or conflicts of interest to disclose.

Data availability

Data is provided within the manuscript or supplementary files.

Declarations

Conflict of interest

The authors declare no competing interests.

Ethical approval

This is a systematic review that does not involve humans or animals in the research. The ethics declaration is not applicable.

Author details

[1]Emergency Department, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, NO. 1 Shuai Fu Yuan,Dongcheng District, Beijing 100730, China. [2]Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences and School of Basic Medicine, Peking Union Medical College, Bejing, China.

Received: 17 October 2024 Accepted: 18 January 2025
Published online: 03 February 2025

References

1. Thomas CP, Ryan M, Chapman JD, Stason WB, Tompkins CP, Suaya JA, Polsky D, Mannino DM, Shepard DS. Incidence and cost of pneumonia in medicare beneficiaries. Chest. 2012;142(4):973–81.
2. Li G, Huang J, Li Y, Feng J. The value of combined radial endobronchial ultrasound-guided transbronchial lung biopsy and metagenomic next-generation sequencing for peripheral pulmonary infectious lesions. Can Respir J. 2020;2020:2367505.
3. Vincent JL, Rello J, Marshall J, et al. International study of the prevalence and outcomes of infection in intensive care units. JAMA. 2009;302(21):2323–9.
4. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009–2014. JAMA. 2017;318(13):1241–9.
5. Postma DF, van Werkhoven CH, van Elden LJ, Thijsen SF, Hoepelman AI, Kluytmans JA, Boersma WG, Compaijen CJ, van der Wall E, Prins JM, et al. Antibiotic treatment strategies for community-acquired pneumonia in adults. N Engl J Med. 2015;372(14):1312–23.
6. Liu VX, Fielding-Singh V, Greene JD, Baker JM, Iwashyna TJ, Bhattacharya J, Escobar GJ. The timing of early antibiotics and hospital mortality in sepsis. Am J Respir Crit Care Med. 2017;196(7):856–63.
7. Cillóniz C, Torres A, Niederman MS: Management of pneumonia in critically ill patients. BMJ 2021;375.
8. Clinical and Laboratory Standards Institute. M100 Performance Standards for Antimicrobial Susceptibility Testing. 30 ed. Wayne, PA: CLSI; 2020.
9. Church DL, Cerutti L, Gürtler A, Griener T, Zelazny A, Emler S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. Clin Microbiol Rev. 2020;33(4):e00053-e119.
10. Vandenberg O, Durand G, Hallin M, Diefenbach A, Gant V, Murray P, et al. Consolidation of clinical microbiology laboratories and introduction of transformative technologies. Clin Microbiol Rev. 2020;33(2):e00057-e119.
11. Chiu CY, Miller SA. Clinical metagenomics. Nat Rev Genet. 2019;20(6):341–55.
12. Miller SC, Rodino C, Miller KG. Point-counterpoint: should we be performing metagenomic next-generation sequencing for infectious disease diagnosis in the clinical laboratory? J Clin Microbiol. 2020;58:01739–11719.
13. Qu J, Zhang J, Chen Y, Huang Y, Xie Y, Zhou M, Li Y, Shi D, Xu J, Wang Q, et al. Aetiology of severe community acquired pneumonia in adults identified by combined detection methods: a multi-centre prospective study in China. Emerg Microbes Infections. 2022;11(1):556–66.
14. Sun L, Zhang S, Yang Z, Yang F, Wang Z, Li H, Li Y, Sun T. Clinical application and influencing factor analysis of metagenomic next-generation sequencing (mNGS) in ICU patients with sepsis. Front Cell Infect Microbiol. 2022;13(12): 905132. https://doi.org/10.3389/fcimb.2022.905132.
15. Shi R, Wang Y, Zhou S, Zhang Y, Zheng S, Zhang D, Du X, Gu W, Xu Y, Zhu C. Metagenomic next-generation sequencing for detecting lower respiratory tract infections in sputum and bronchoalveolar lavage fluid samples from children. Front Cell Infect Microbiol. 2023;17(13):1228631. https://doi.org/10.3389/fcimb.2023.1228631.
16. Zhong J, Liu Y, Luo N, Wei Q, Qisi S, Zou J, Xiaozhong W, Huang X, Jiang Y, Liang L, Li H, Lin J. Metagenomic next-generation sequencing for rapid detection of pulmonary infection in patients with acquired immunodeficiency syndrome. Ann Clin Microbiol Antimicrob. 2023. https://doi.org/10.1186/s12941-023-00608-9.
17. Zhao X, Duan MX, Lu YY, Bai LP, Zhao XY: Short-term prognostic analysis of patients with systemic lupus erythematosus co-infection and

comparison of mNGS and conventional microbiological test results. Frontiers Cell Infect Microbiol 2023;13.
18. Zhang S, Ou J, Tan Y, Yang B, Wu Y, Liu L. Metagenomic next-generation sequencing for pulmonary infections diagnosis in patients with diabetes. BMC Pulm Med. 2023;23(1):142.
19. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
20. Liu Y, Qin S, Lan C, Huang Q, Zhang P, Cao W. Effectiveness of metagenomic next-generation sequencing in the diagnosis of infectious diseases: a systematic review and meta-analysis. Int J Infect Dis. 2024;142: 106996.
21. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data—which coefficients and confidence intervals are appropriate? BMC Med Res Methodol. 2016;16:93.
22. He S, Liu H, Hu X, Zhao J, Liang J, Zhang X, Chen J, Zeng H, Sun G. Exploring the clinical and diagnostic value of metagenomic next-generation sequencing for urinary tract infection: a systematic review and meta-analysis. BMC Infect Dis. 2024;24(1):1000.
23. Chen Y, Wang J, Niu T. Clinical and diagnostic values of metagenomic next-generation sequencing for infection in hematology patients: a systematic review and meta-analysis. BMC Infect Dis. 2024;24(1):167.
24. Liu J, Zhang Q, Dong YQ, Yin J, Qiu YQ. Diagnostic accuracy of metagenomic next-generation sequencing in diagnosing infectious diseases: a meta-analysis. Sci Rep. 2022;12(1):21032.
25. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–82.
26. Shi W, Zhu S. The application of metagenomic next-generation sequencing in detection of pathogen in bronchoalveolar lavage fluid and sputum samples of patients with pulmonary infection. Comput Math Methods Med. 2021;8(2021):7238495.
27. Ju CR, Lian QY, Guan WJ, Chen A, Zhang JH, Xu X, Chen RC, Li SY, He JX. Metagenomic next-generation sequencing for diagnosing infections in lung transplant recipients: a retrospective study. Transpl Int. 2022;10(35):10265.
28. Shi Y, Wu J, Liu T, Yue L, Liu Y, Gu Y, Qi Y. Analysis of metagenomic next-generation sequencing results of 25 pus samples. Infect Drug Resist. 2022;8(15):6515–24.
29. Wang JZ, Yuan D, Yang XH, Sun CH, Hou LL, Zhang Y, Gao YX. Etiology of lower respiratory tract in pneumonia based on metagenomic next-generation sequencing: a retrospective study. Front Cell Infect Microbiol. 2024;9(13):1291980.
30. Zhang L, Fang K, Ren H, Fan S, Wang J, Guan H. Comparison of the diagnostic significance of cerebrospinal fluid metagenomic next-generation sequencing copy number variation analysis and cytology in leptomeningeal malignancy. BMC Neurol. 2024;24(1):223.
31. Zhang X, Ye J, Wang L, Zhang L, Wang L, Jin H. Nanopore sequencing technology: a reliable method for pathogen diagnosis in elderly patients with community-acquired pneumonia. Infect Drug Resist. 2024;24(17):3659–67.
32. Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med. 2017;43:304–77.
33. Liu D, Huang SY, Sun JH, Zhang HC, Cai QL, Gao C, Li L, Cao J, Xu F, Zhou Y, Guan CX, Jin SW, Deng J, Fang XM, Jiang JX, Zeng L. Sepsis-induced immunosuppression: mechanisms, diagnosis and current treatment options. Mil Med Res. 2022;9(1):56.
34. Pan TJ, Luo WW, Zhang SS, Xie JY, Xu Z, Zhong YY, Zou XF, Gong HJ, Ye ML. The clinical application value of multi-site mNGS detection of patients with sepsis in intensive care units. BMC Infect Dis. 2024;24(1):920.
35. Song J, Liu S, Xie Y, Zhang C, Xu C. Diagnostic value of DNA or RNA-based metagenomic next-generation sequencing in lower respiratory tract infections. Heliyon. 2024;10(9): e30712.
36. Yuan L, Zhu XY, Lai LM, Chen Q, Liu Y, Zhao R. Clinical application and evaluation of metagenomic next-generation sequencing in pathogen detection for suspected central nervous system infections. Sci Rep. 2024;14(1):16961.
37. Liu H, Ran Q, Ma J, Zhang J, Tan N, Xi L, Li X, Zhang J, Lu S. Retrospective clinical and microbiologic analysis of metagenomic next-generation sequencing in the microbiological diagnosis of cutaneous infectious granulomas. Ann Clin Microbiol Antimicrob. 2024;23(1):84.

Publisher's Note