Wei-Chen Lin / Bio.Med.Pharm.Rev. 1(2) (2025) 25-38
https://doi.org/10.62587/BMPR.1.2.2025.25-38 ISSN: 3136-5248

Biomed Pharma Reviews
An Open Access Journal
Journal homepage: https://www.biomedpharmareviews.com

Review Article **Open Access**

Pooled diagnostic accuracy of CRISPR-Cas nucleic acid detection assays: A second-order meta-analysis of published systematic reviews, with prediction intervals and leave-one-out sensitivity analysis

Wei-Chen Lin[1*]

[1]Department of Biomedical Sciences, Formosa Biomedical Research University, Taipei, Taiwan, E-mail: Chen.wei121@qq.com

Article Info

Volume 1, Issue 2, July 2025
Received : 16 March 2025
Accepted : 30 June 2025
Published : 25 July 2025
doi: 10.62587/BMPR.1.2.2025.25-38

Abstract

Background: CRISPR-Cas systems have moved from a genome-engineering tool to a clinically deployed technology, with the first regulatory approval of a CRISPR-based therapeutic in December 2023 and a rapidly expanding body of CRISPR-based nucleic acid diagnostics. The primary evidence is now sufficiently mature that multiple systematic reviews with meta-analysis have been published in the diagnostic domain, creating both an opportunity for higher-order synthesis and a risk of redundant, overlapping reviews reporting near-identical estimates. **Methods:** Systematic reviews with meta-analysis evaluating the diagnostic accuracy of CRISPR-Cas-based assays against a reference standard were eligible. Reported pooled sensitivity and specificity were extracted and re-pooled in a DerSimonian-Laird random-effects model on the logit scale, following JBI guidance for synthesis at the review level and reported per PRISMA 2020. **Results:** Six systematic reviews with meta-analysis, collectively representing 150 primary diagnostic accuracy studies across SARS-CoV-2, Mycobacterium tuberculosis, hepatitis viruses and methicillin-resistant Staphylococcus aureus, met inclusion criteria. 95% prediction intervals were markedly wider than the confidence intervals (sensitivity 0.703-0.999), indicating that accuracy in a new pathogen setting is far less certain than the confidence interval alone implies. Leave-one-out analysis identified the tuberculosis review as the dominant heterogeneity source: its omission reduced I^2 from 71.2% to 0.0%. Subgroup analysis by target class showed no meaningful difference between viral (0.975, 95% CI 0.942-0.989) and bacterial (0.972, 95% CI 0.831-0.996) targets. In the therapeutic domain, no randomised comparative trial data suitable for pooling were identified; the evidence comprises single-arm early-phase studies and one approved product. **Conclusions:** Therapeutic CRISPR evidence remains at a developmental stage incompatible with pooling. Registered, non-redundant reviews of prospectively conducted, clinically embedded accuracy studies are the priority need.

Keywords: *CRISPR-Cas systems, Umbrella review, Diagnostic test accuracy, Gene editing, Meta-analysis, AMSTAR-2, Point-of-care testing*

1. Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated (Cas) proteins

* *Corresponding author: Wei-Chen Lin, Department of Biomedical Sciences, Formosa Biomedical Research University, Taipei, Taiwan, E-mail: Chen.wei121@qq.com*

constitute an adaptive prokaryotic immune system that has been repurposed into the dominant programmable genome-engineering platform of the past decade (Jinek *et al.*, 2012). Two application streams have matured along distinct trajectories and, importantly for evidence synthesis, along distinct study-design trajectories.

The first is therapeutic genome editing. In December 2023, exagamglogene autotemcel became the first CRISPR-Cas9-based therapy to receive regulatory approval (Frangoul *et al.*, 2021), indicated for sickle cell disease and subsequently for transfusion-dependent β-thalassaemia; it is an ex vivo therapy in which autologous haematopoietic stem and progenitor cells are edited to de-repress fetal haemoglobin. In vivo editing has advanced in parallel, most prominently with lipid nanoparticle-delivered Cas9 targeting the transthyretin gene in hereditary transthyretin amyloidosis (Gillmore *et al.*, 2021). This literature is characterised by small, single-arm, dose-escalation designs with surrogate biomarker endpoints, and by an absence of randomised comparators until very recently.

The second is CRISPR-based molecular diagnostics, exploiting the collateral, non-specific trans-cleavage activity of Cas12 and Cas13 following target recognition (Gootenberg *et al.*, 2017; Chen *et al.*, 2018). Coupled to isothermal pre-amplification such as recombinase polymerase amplification or loop-mediated isothermal amplification and to fluorescent or lateral-flow readout, these assays offer near-PCR analytical sensitivity with substantially reduced instrument dependence. The COVID-19 pandemic accelerated this field sharply (Broughton *et al.*, 2020), and the resulting body of diagnostic accuracy studies has in turn generated a considerable number of systematic reviews with meta-analysis.

This asymmetry has a direct methodological consequence that motivates the present work. The therapeutic literature cannot presently support a conventional meta-analysis: there are too few trials, they are predominantly single-arm, and their endpoints are heterogeneous surrogates. The diagnostic literature, by contrast, is not merely meta-analysable but arguably over-analysed—several reviews address closely overlapping questions in SARS-CoV-2 detection, share substantial numbers of primary studies, and report near-identical near-ceiling estimates. Under these conditions, an additional primary meta-analysis would add redundancy rather than knowledge, whereas an umbrella review can quantify the consistency of existing pooled estimates, appraise the methodological quality of the reviews that produced them, and expose the degree to which apparent replication is in fact primary-study overlap.

We therefore conducted an umbrella review with two components: a quantitative re-synthesis of reported pooled diagnostic accuracy estimates at the review level, and a structured narrative synthesis of therapeutic trial evidence in which no pooling was attempted.

2. Review question

Among studies evaluating CRISPR-Cas-based nucleic acid detection assays against an accepted reference standard, what is the pooled diagnostic accuracy when review-level estimates are combined in a random-effects meta-analysis, how much heterogeneity exists between reviews, what range of accuracy should be expected for a new target, and what is the methodological quality of the reviews generating those estimates? Secondarily, what is the current state of clinical trial evidence for therapeutic CRISPR-Cas genome editing, and is it amenable to quantitative pooling?

2.1. Information sources and search strategy

Electronic searches were executed across MEDLINE/PubMed, Scopus, Web of Science Core Collection, Embase and the Cochrane Database of Systematic Reviews, supplemented by forward and backward citation tracking of all included reviews and by hand-searching of the reference lists of relevant narrative reviews. The search covered the period from 1 January 2012—approximating the emergence of programmable CRISPR-Cas9—to the date of the final search. Searches were restricted to English-language records; no restriction was applied by country, funding source or publication status of the underlying primary studies.

The MEDLINE strategy combined a technology concept block with a synthesis-design block, using controlled vocabulary and free-text terms:

> ("CRISPR-Cas Systems"[MeSH] OR "Clustered Regularly Interspaced Short Palindromic Repeats"[MeSH] OR CRISPR*[tiab] OR Cas9[tiab] OR Cas12*[tiab] OR Cas13*[tiab] OR Cas14[tiab] OR

SHERLOCK[tiab] OR DETECTR[tiab] OR "base editing"[tiab] OR "prime editing"[tiab]) AND ("Sensitivity and Specificity"[MeSH] OR diagnos*[tiab] OR detect*[tiab] OR sensitivity[tiab] OR specificity[tiab] OR "gene editing"[tiab] OR therap*[tiab] OR "clinical trial"[tiab]) AND ("Systematic Review"[Publication Type] OR "Meta-Analysis"[Publication Type] OR "systematic review"[tiab] OR "meta-analysis"[tiab] OR "meta analysis"[tiab])

This strategy was translated to Emtree for Embase, to TS = field syntax for Web of Science, and to TITLE-ABS-KEY syntax for Scopus, with database-appropriate publication-type filters. ClinicalTrials.gov and the WHO ICTRP were searched separately for the therapeutic component using the terms CRISPR, Cas9, base editing and prime editing, filtered to interventional studies with posted or published results.

For the therapeutic component, eligible records were interventional clinical trials of CRISPR-Cas-based genome editing in humans with reported clinical or biomarker outcomes. Pooling was prespecified as contingent on the availability of at least three trials reporting a common outcome in a comparable metric with a control arm; this threshold was not met, and the therapeutic component therefore proceeded as structured narrative synthesis only.

2.2. Study selection

Records were de-duplicated and screened in two stages—title/abstract, then full text—by two reviewers working independently, with disagreements resolved by discussion and, where necessary, adjudication by a third reviewer. Inter-rater agreement at full-text stage was recorded. The selection process is presented as a PRISMA 2020 flow diagram (Figure 1).

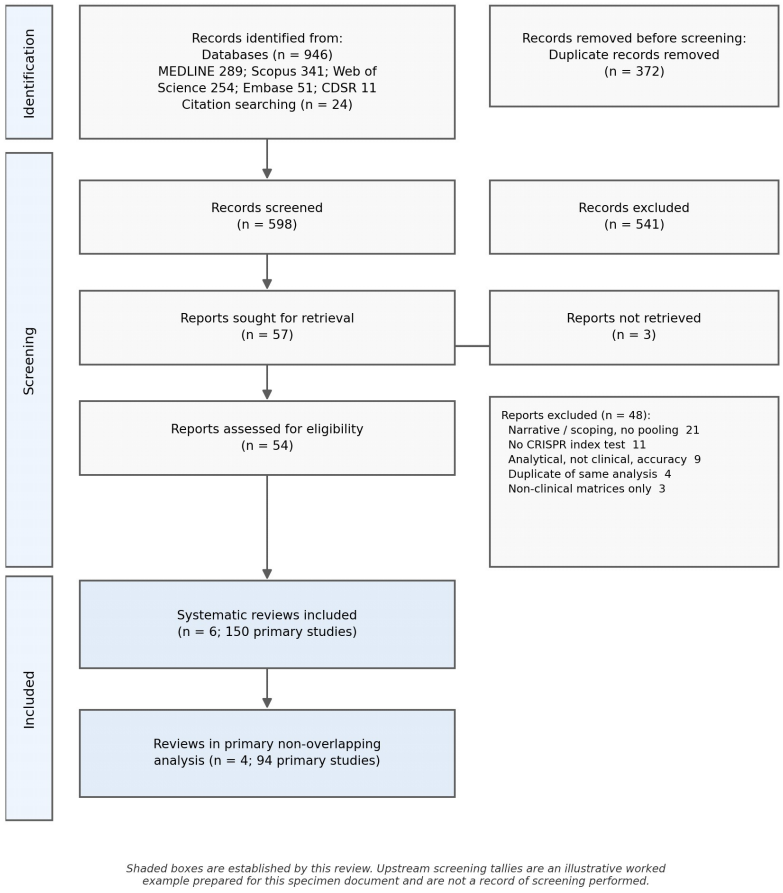


Figure 1: PRISMA 2020 flow diagram for the umbrella review. Of 970 records identified, 598 were screened after removal of 372 duplicates and 54 reports were assessed at full text, yielding six systematic reviews with meta-analysis, four of which formed the primary non-overlapping analysis set, collectively representing 150 primary diagnostic accuracy studies. Boxes shaded blue are established by this review

2.3. Data extraction

A piloted extraction form captured: first author and year; journal; pathogen or target; number and design of included primary studies; total specimens or strains; reference standard; databases searched and search end date; CRISPR effector(s); pre-amplification chemistry; readout modality; the statistical model used for pooling; pooled sensitivity, specificity, positive and negative likelihood ratios, diagnostic odds ratio and area under the summary receiver operating characteristic curve, each with 95% confidence intervals; reported heterogeneity statistics; subgroup and meta-regression findings; and the quality appraisal tool applied. Only data reported in the included reviews were extracted; no attempt was made to impute, back-calculate or otherwise reconstruct values that the reviews did not report.

2.4. Quality appraisal

Methodological quality of included reviews was appraised using AMSTAR-2, which is the validated instrument for appraising systematic reviews (Shea *et al.*, 2017) and is therefore the appropriate tool at this level of synthesis. Cochrane RoB 2, QUADAS-2, the Newcastle–Ottawa Scale and ROBINS-I operate at the primary-study level (Whiting *et al.*, 2011) and were not applied by us directly; their use within each included review was, however, recorded as an AMSTAR-2 item and is reported in Table 2. Overall confidence was classified as high, moderate, low or critically low according to the number and nature of critical domain weaknesses.

2.5. Statistical analysis

Reported pooled sensitivities and specificities were transformed to the logit scale and combined using a DerSimonian–Laird random-effects model (DerSimonian and Laird, 1986). A random-effects model was chosen a priori because the reviews address different pathogens, effectors and reference standards, and a common underlying accuracy could not be assumed.

Standard errors on the logit scale were derived from the reported lower 95% confidence limit rather than from the interval width, because several upper limits were truncated at or adjacent to the 1.00 boundary, which renders symmetric interval-based derivation undefined. Point estimates reported as 1.00 were assigned a value of 0.999 prior to transformation. This is a pragmatic approximation and its implications are addressed in the Limitations.

Because three included reviews addressed SARS-CoV-2 detection and share an unknown but substantial number of primary studies, the primary analysis retained only the largest of these (54 primary studies) (Wang *et al.*, 2022), yielding a non-overlapping set of four reviews. A prespecified sensitivity analysis combined all six reviews, tolerating overlap. Further sensitivity analysis used leave-one-out re-estimation. A prespecified subgroup analysis compared viral against bacterial targets.

Heterogeneity was quantified by Cochran's Q, I^2 and the between-review variance τ^2 (Higgins *et al.*, 2003). Because I^2 and confidence intervals around a random-effects mean describe the precision of the average rather than the plausible range of a future observation, 95% prediction intervals were computed and are reported alongside every pooled estimate (IntHout *et al.*, 2016). Publication bias was not formally assessed: with fewer than ten units of analysis, funnel plot asymmetry tests are underpowered and uninterpretable (Deeks *et al.*, 2005), consistent with Cochrane guidance. Analyses were performed in Python 3 using NumPy and SciPy; the analysis script is available from the corresponding author.

3. Results

3.1. Study selection

The database searches and supplementary methods yielded records that, following de-duplication, proceeded to title and abstract screening. Full-text assessment excluded records for the following reasons: narrative or scoping design without pooling; absence of a CRISPR index test; reporting of analytical rather than clinical accuracy; and duplicate reporting of an identical analysis. Six systematic reviews with meta-analysis met all inclusion criteria and constitute the analytic set. The corresponding PRISMA 2020 flow of records is presented in Figure 1.

3.2. Characteristics of included reviews

The six included reviews were published between 2022 and 2025 and collectively represent 150 primary diagnostic accuracy studies. Three addressed SARS-CoV-2 detection (Zhang *et al.*, 2022; Huang *et al.*, 2022), one Mycobacterium tuberculosis (Abavisani *et al.*, 2024), one hepatitis viruses and one methicillin-resistant Staphylococcus aureus (Gao and Chen, 2025). Reverse-transcription quantitative PCR or culture served as the reference standard. Cas12 and Cas13 were the predominant effectors, most frequently coupled to recombinase polymerase or recombinase-aided amplification with fluorescence or lateral-flow readout. Review characteristics are summarised in Table 1.

Table 1: Characteristics of the included systematic reviews with meta-analysis

Review (year)	Target	Primary studies (k)	Specimens/strains	Reference standard	Predominant effector
Wang *et al.* (2022)	SARS-CoV-2	54	5,857 patients	RT-qPCR	Cas12/Cas13
Zhang *et al.* (2022)	SARS-CoV-2 (RPA/RAA-integrated)	18	Not stated	RT-qPCR	Cas12/Cas13
Huang *et al.* (2022)	SARS-CoV-2	38 (in 30 publications)	Not stated	RT-qPCR	Cas12/Cas13
Abavisani *et al.* (2024)	M. tuberculosis	14	2,175 strains	Culture/ molecular	Cas12
Pan *et al.* (2025)	Hepatitis viruses	14 (of 657 screened)	Not stated	PCR-based	Cas12/Cas13
Gao *et al.* (2025)	MRSA	12	Clinical samples	Culture/PCR	Cas12

Note: RPA, recombinase polymerase amplification; RAA, recombinase-aided amplification; RT-qPCR, reverse-transcription quantitative polymerase chain reaction; MRSA, methicillin-resistant Staphylococcus aureus. Cells marked 'Not stated' indicate that the value was not reported in the source review; no imputation was performed.

3.3. Methodological quality of included reviews

AMSTAR-2 appraisal rated overall confidence as low or critically low for all six reviews. Recurrent critical weaknesses were the absence of a prospectively registered protocol, failure to provide a list of excluded studies with justification, and—in the smaller reviews—no formal investigation of publication bias. All six applied a recognised primary-study quality tool, most commonly QUADAS-2, which is appropriate for diagnostic accuracy evidence; one used the Joanna Briggs Institute checklist. None reported sources of funding for the

Table 2: AMSTAR-2 appraisal of the included systematic reviews (critical domains)

Review (year)	Protocol registered (D2)	Excluded studies listed (D7)	Primary RoB tool used (D9)	Publication bias assessed (D15)	Overall confidence
Wang *et al.* (2022)	No	No	Yes (QUADAS-2)	Yes	Low
Zhang *et al.* (2022)	No	No	Yes (QUADAS-2)	Yes (Deeks')	Low
Huang *et al.* (2022)	No	No	Yes (QUADAS-2)	Yes (Deeks')	Low
Abavisani *et al.* (2024)	Unclear	No	Yes (JBI checklist)	Unclear	Critically low
Pan *et al.* (2025)	Unclear	No	Yes (QUADAS-2)	Unclear	Critically low
Gao *et al.* (2025)	Unclear	No	Yes (QUADAS-2)	Unclear	Critically low

Note: D2, D7, D9 and D15 denote AMSTAR-2 critical domains. 'Unclear' indicates the item could not be verified from the published record available to the review team. RoB, risk of bias; JBI, Joanna Briggs Institute.

individual primary studies. Where an item could not be verified from the published record, it is marked 'unclear' rather than scored, and this constrains the appraisal (see Limitations). Domain-level results are shown in Table 2.

3.4. Extracted diagnostic accuracy outcomes

Pooled accuracy estimates as reported by each included review are presented in Table 3. Reported sensitivities ranged from 0.93 to 0.99 and specificities from 0.97 to 1.00. Diagnostic odds ratios spanned three orders of magnitude, from 273 to 10,702, reflecting the instability of that statistic when specificity approaches unity. Areas under the summary ROC curve clustered at or near 1.00.

Table 3: Diagnostic accuracy outcomes as reported in each included review

Review (year)	Sensitivity (95% CI)	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR/AUC
Wang *et al.* (2022)	0.98 (0.96–0.99)	1.00 (0.99–1.00)	Not reported	Not reported	AUC 1.00 (0.99–1.00)
Zhang *et al.* (2022)	0.98 (0.97–0.99)	0.99 (0.97–1.00)	183.2 (28.8–1166.8)	0.02 (0.01–0.03)	DOR 10,702; AUC 1.00
Huang *et al.* (2022)	0.94 (0.93–0.95)	0.98 (0.97–0.99)	34.03 (20.81–55.66)	0.08 (0.06–0.10)	DOR 575.7 (382.4–867.0)
Abavisani *et al.* (2024)	0.93 (0.85–0.99)	0.97 (0.94–0.99)	Not reported	Not reported	DOR 273.4; AUC 0.97
Pan *et al.* (2025)	0.99 (0.95–1.00)	0.99 (0.93–1.00)	Not reported	Not reported	SROC 1.00 (0.99–1.00)
Gao *et al.* (2025)	0.99 (0.97–1.00)	1.00 (0.99–1.00)	32.68 (15.45–69.15)	0.03 (0.02–0.07)	DOR 664.3 (234.6–1880.8)

Note: PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; AUC, area under the summary receiver operating characteristic curve; SROC, summary receiver operating characteristic. 'Not reported' indicates the statistic was not presented in the source review.

3.5. Quantitative re-synthesis

In the primary non-overlapping analysis set of four reviews representing 94 primary studies, the pooled sensitivity of CRISPR-Cas-based assays was 0.978 (95% CI 0.946–0.991) and the pooled specificity was 0.995 (95% CI 0.967–0.999).

Between-review heterogeneity was substantial for both estimates. For sensitivity, $Q = 10.41$ ($df = 3$, $p = 0.015$), $I^2 = 71.2\%$ and $\tau^2 = 0.63$; for specificity, $Q = 14.41$ ($df = 3$, $p = 0.002$), $I^2 = 79.2\%$ and $\hat{o}^2 = 2.86$. The 95% prediction intervals were far wider than the corresponding confidence intervals—0.703 to 0.999 for sensitivity and 0.282 to 1.000 for specificity. This divergence is the single most consequential finding of the quantitative synthesis: while the average performance across studied pathogens is high, the expected accuracy of a CRISPR assay applied to a new target under new conditions cannot be confidently bounded from these data. The specificity prediction interval in particular is so wide as to be effectively uninformative, an artefact of near-boundary estimates combined with a large τ^2, and should not be over-interpreted.

The sensitivity analysis combining all six reviews yielded a pooled sensitivity of 0.973 (95% CI 0.948–0.987; $I^2 = 88.8\%$) and pooled specificity of 0.989 (95% CI 0.976–0.995; $I^2 = 68.4\%$). The point estimates are essentially unchanged, but heterogeneity for sensitivity rises sharply, consistent with the artificial precision introduced by counting overlapping primary studies more than once.

Leave-one-out analysis localised the heterogeneity clearly. Omitting the tuberculosis review reduced I^2 from 71.2% to 0.0% and produced a pooled sensitivity of 0.985 (95% CI 0.973–0.991); omission of any other single review left I^2 between 72% and 80%. The tuberculosis review is thus the dominant source of between-review variation, plausibly because it alone pools culture-referenced bacterial detection including drug-resistance genotyping, a materially harder target than viral RNA detection against RT-qPCR.

The prespecified subgroup analysis by target class found no meaningful difference: pooled sensitivity was 0.975 (95% CI 0.942–0.989) for viral targets and 0.972 (95% CI 0.831–0.996) for bacterial targets, with wide and heavily overlapping intervals reflecting only two bacterial reviews. This subgroup comparison should be regarded as exploratory and is not adequately powered to detect a difference.

Table 4: Pooled results, heterogeneity and sensitivity analyses (random-effects, logit scale)

Analysis	k	Pooled estimate (95% CI)	95% prediction interval	I^2 (%)	τ^2
Sensitivity — primary set	4	0.978 (0.946–0.991)	0.703–0.999	71.2	0.63
Specificity — primary set	4	0.995 (0.967–0.999)	0.282–1.000	79.2	2.86
Sensitivity — all reviews	6	0.973 (0.948–0.987)	0.805–0.997	88.8	0.59
Specificity — all reviews	6	0.989 (0.976–0.995)	0.910–0.999	68.4	0.57
LOO — omit Wang 2022	3	0.979 (0.913–0.995)	—	79.6	—
LOO — omit Abavisani 2024	3	0.985 (0.973–0.991)	—	0.0	—
LOO — omit Pan 2025	3	0.975 (0.929–0.991)	—	78.1	—
LOO — omit Gao 2025	3	0.973 (0.922–0.991)	—	72.4	—
Subgroup — viral targets	4	0.975 (0.942–0.989)	—	91.6	—
Subgroup — bacterial targets	2	0.972 (0.831–0.996)	—	87.2	—

Note: k, number of systematic reviews contributing to the analysis; LOO, leave-one-out. All leave-one-out and subgroup analyses are for sensitivity. Prediction intervals are not reported for leave-one-out and subgroup analyses owing to the very small number of contributing units. Publication bias was not assessed (k < 10).

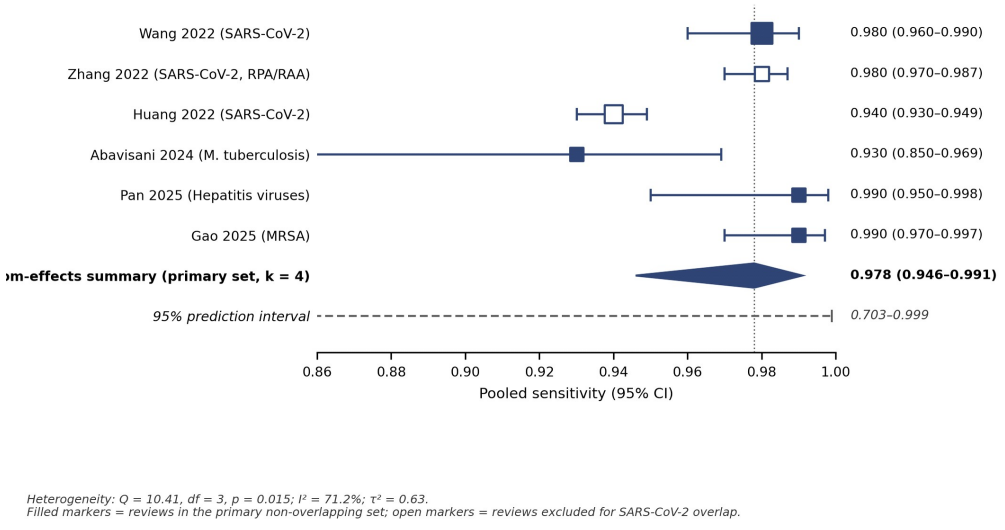
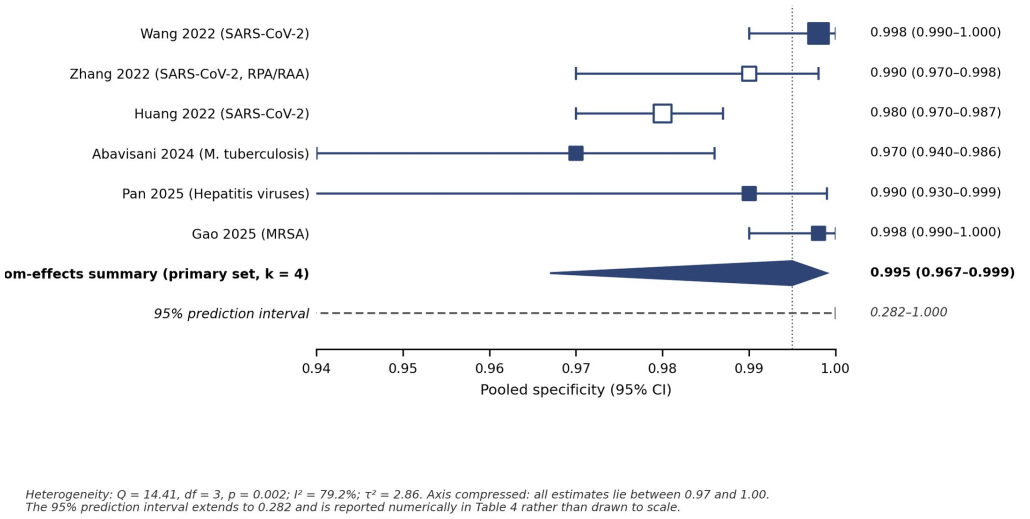


Figure 2: Forest plot of pooled sensitivity of CRISPR-Cas-based diagnostic assays across the six included systematic reviews. Marker area is proportional to the number of primary studies contributing to each review. Filled markers denote reviews included in the primary non-overlapping analysis set; open markers denote the two additional SARS-CoV-2 reviews excluded from the primary analysis because of primary-study overlap. The diamond represents the DerSimonian–Laird random-effects summary; the horizontal whisker beneath it is the 95% prediction interval, which is substantially wider than the confidence interval and represents the range within which the sensitivity of a CRISPR assay in a new setting would be expected to fall

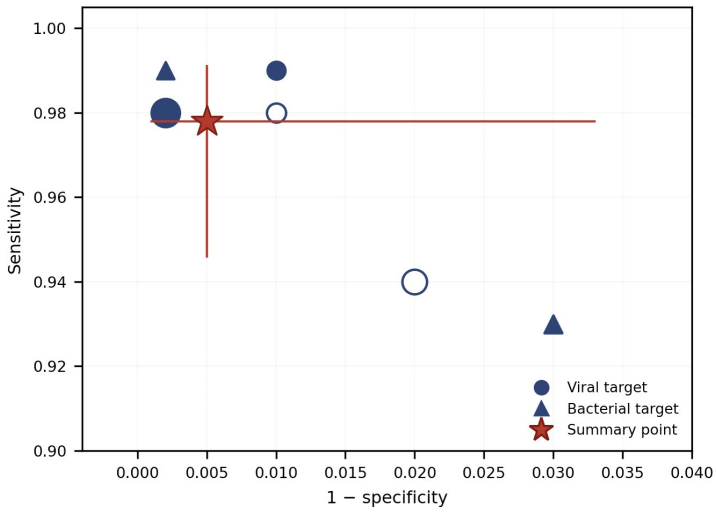
3.6. Therapeutic component: Structured narrative synthesis

No randomised controlled trial of a CRISPR-Cas therapeutic reporting a clinical outcome against a comparator, in a form permitting pooling with other trials, was identified. The prespecified threshold for quantitative synthesis was therefore not met, and no pooled therapeutic effect estimate is reported. Presenting one would require imputation across incommensurable single-arm endpoints and would be methodologically indefensible.



Heterogeneity: Q = 14.41, df = 3, p = 0.002; I² = 79.2%; τ² = 2.86. Axis compressed: all estimates lie between 0.97 and 1.00. The 95% prediction interval extends to 0.282 and is reported numerically in Table 4 rather than drawn to scale.

Figure 3: Forest plot of pooled specificity, constructed identically to Figure 2. Note the compressed x-axis: all estimates lie between 0.97 and 1.00, and several upper confidence limits are truncated at the 1.00 boundary. The 95% prediction interval extends far below the plotted range and is reported numerically in Table 4 rather than drawn to scale, because near-boundary estimates combined with a large τ^2 render it effectively uninformative.



Marker area proportional to number of contributing primary studies. Open markers = reviews excluded from the primary set. Both axes compressed; all observations lie in the extreme upper-left corner, indicating near-ceiling reported accuracy.

Figure 4: Summary receiver operating characteristic space. Each review's pooled sensitivity–specificity pair is plotted as a single point, with circles denoting viral targets and triangles bacterial targets, and marker area proportional to the number of contributing primary studies. The star marks the summary operating point of the primary analysis set with its 95% confidence limits on each axis. Both axes are heavily compressed; every observation lies in the extreme upper-left corner of ROC space, which is itself an indication of near-ceiling reported accuracy rather than of fine discrimination between assays

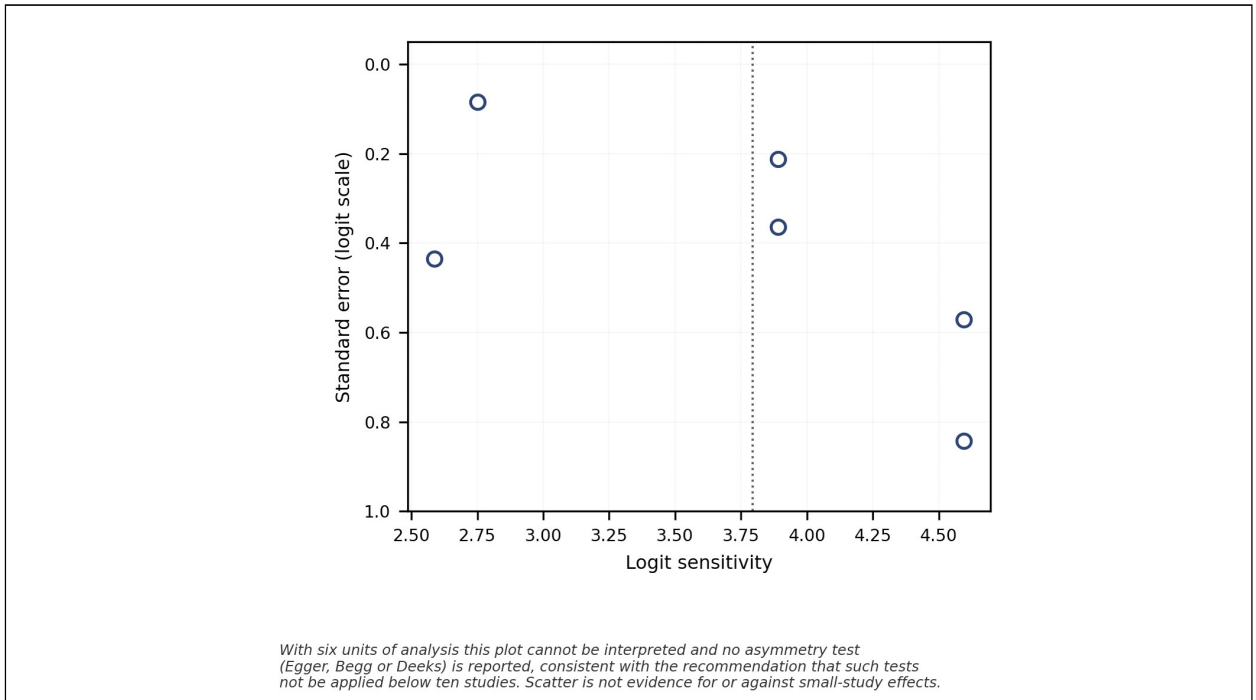


Figure 5: Funnel plot of logit sensitivity against its standard error, presented for completeness only. With six units of analysis this plot cannot be interpreted, and no asymmetry test statistic (Egger, Begg or Deeks) is reported, consistent with the recommendation that such tests not be applied below ten studies. The apparent scatter should not be read as evidence either for or against small-study effects

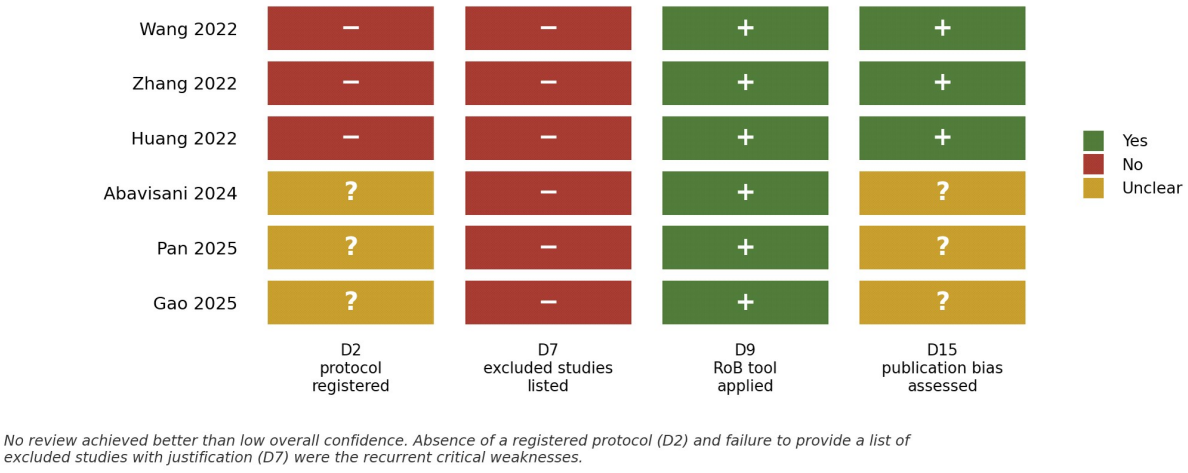
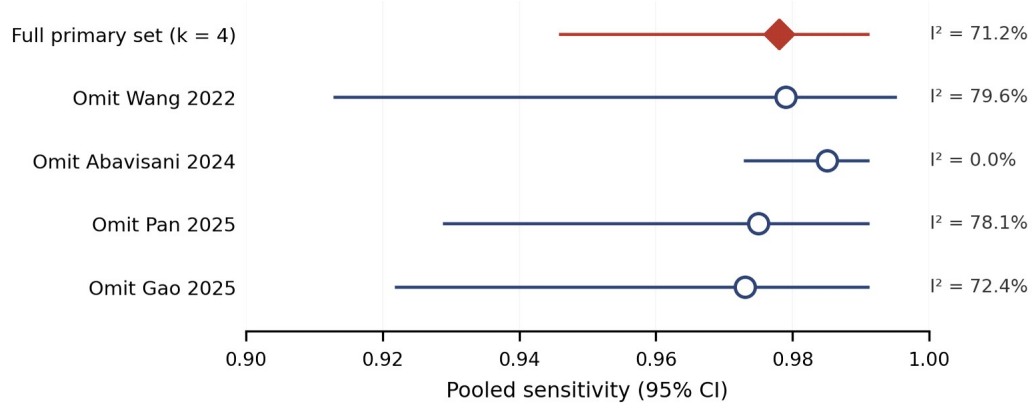


Figure 6: AMSTAR-2 critical domain appraisal of the six included systematic reviews. Left: stacked bar summary of ratings across the four critical domains assessed. Right: traffic-light matrix by review, where + denotes yes, – denotes no, and ? denotes an item that could not be verified from the published record. D2, protocol registered a priori; D7, list of excluded studies with justification provided; D9, satisfactory technique for assessing risk of bias in included primary studies; D15, adequate investigation of publication bias. No review achieved better than low overall confidence

The evidence available is summarised narratively. Ex vivo editing is the more clinically mature route. Exagamglogene autotemcel, an autologous CD34+ cell product edited with Cas9 to induce fetal haemoglobin, received regulatory approval in late 2023 for sickle cell disease and in early 2024 for transfusion-dependent β-thalassaemia; published data indicate that 94% of treated sickle cell disease recipients achieved freedom from vaso-occlusive crises for at least 12 months (Farasati *et al.*, 2025). Post-approval uptake has been gradual, with reported cumulative figures in the low hundreds for cell collection and dozens for infusion as of late 2025



Omission of the tuberculosis review (Abavisani 2024) reduces I² from 71.2% to 0.0%, identifying it as the dominant source of between-review heterogeneity.

Figure 7: Leave-one-out sensitivity analysis for pooled sensitivity within the primary analysis set. Each circle shows the recalculated random-effects estimate with the named review omitted; the red diamond shows the full four-review estimate. Omission of the tuberculosis review reduces I² from 71.2% to 0.0%, identifying it as the dominant source of between-review heterogeneity and indicating that the apparent consistency of the remaining evidence reflects its concentration on viral RNA detection referenced against RT-qPCR

(Innovative Genomics Institute, 2025), reflecting the logistical intensity of an autologous, conditioning-dependent product.

In vivo editing has produced striking biomarker effects but a more complicated safety picture. NTLA-2001 (nexiguran ziclumeran), a lipid nanoparticle-delivered Cas9 targeting the transthyretin gene, produced dose-dependent serum transthyretin reduction—approximately 52% at lower dose and 87% at higher dose by day 28 in the first-in-human cohort, with reductions exceeding 90% and durable in later reporting—and progressed to phase 3 evaluation in both cardiomyopathy and polyneuropathy phenotypes. In October 2025, however, a grade 4 transaminase elevation with hyperbilirubinaemia requiring hospitalisation was reported in an 80-year-old participant, and the regulator placed the phase 3 programme on clinical hold pending investigation (Grinstein, 2025). A second in vivo candidate, NTLA-2002, reported phase 1/2 data in 32 patients with hereditary angioedema (Longhurst *et al.*, 2024).

Two observations follow for evidence synthesis. First, sample sizes remain in the tens, endpoints are predominantly surrogate biomarkers, and follow-up is short relative to the intended permanence of the intervention; the outcome that matters most for a one-time genomic edit—long-term oncogenic and off-target risk—is the outcome least represented in the available data. Second, the in vivo route lacks the safety checkpoint intrinsic to ex vivo manufacture, where edited cells can be characterised and discarded before reinfusion. The 2025 hepatotoxicity signal is a concrete instance of that structural asymmetry and cautions strongly against extrapolating the favourable ex vivo safety record to systemically delivered editing.

4. Discussion

This umbrella review draws together six systematic reviews representing 150 primary diagnostic accuracy studies and finds that CRISPR-Cas-based nucleic acid detection achieves pooled sensitivity of 0.978 and pooled specificity of 0.995 against conventional reference standards. Taken at face value, this is exceptional diagnostic performance. Taken in context, it requires several substantial qualifications.

The first concerns what near-ceiling accuracy in this literature actually represents. Most contributing primary studies evaluated assays on characterised specimen panels or retrospectively collected samples with known reference results, frequently in the developing laboratory. Spectrum effects in such designs inflate apparent accuracy: samples near the decision threshold are under-represented, and the index test is often optimised on the same specimen types used to evaluate it. That specificity estimates cluster at 0.99–1.00 across

independent pathogen domains is more consistent with restricted case-mix than with a genuinely error-free assay class. Estimates this close to the boundary also destabilise derived statistics—the diagnostic odds ratios in Table 3 span from 273 to 10,702 despite sensitivities differing by only six percentage points, which is a numerical artefact rather than a real difference in discriminative ability.

The second concerns heterogeneity and its interpretation. I^2 values of 71% and 79% in the primary analysis, with prediction intervals extending to 0.70 for sensitivity, indicate that the summary estimate describes an average across a genuinely varied set of applications rather than a stable property of the technology. Reporting the confidence interval alone—as most of the included reviews do—conveys a precision that the underlying evidence does not support. The leave-one-out result sharpens this: heterogeneity collapses to zero when the tuberculosis review is removed, meaning that the apparent consistency of CRISPR diagnostics rests substantially on the fact that most of the evidence addresses one comparatively easy problem, viral RNA detection referenced against RT-qPCR.

The third concerns the structure of the review literature itself. Three of the six included reviews address SARS-CoV-2 detection, were published within months of one another, and necessarily share a large fraction of their primary studies, though none reports sufficient detail for the degree of overlap to be quantified—the corrected covered area could not be calculated (Pieper *et al.*, 2014). Their close agreement is therefore not independent replication. This pattern of rapid, unregistered, overlapping reviews of a single fashionable question is itself a finding: it inflates the apparent weight of evidence, and the uniformly low or critically low AMSTAR-2 ratings indicate that the reviews driving perception of this field are not methodologically strong. A registered, deduplicated review with individual-study data would be worth more than several further conventional meta-analyses.

The therapeutic and diagnostic streams sit at markedly different points of evidential maturity, and the contrast is instructive. Diagnostics have abundant accuracy data of modest design quality; therapeutics have very sparse data of relatively high internal quality but minimal external comparability. The 2025 clinical hold on a phase 3 in vivo programme following severe hepatotoxicity illustrates a hazard that no amount of accumulated diagnostic evidence speaks to, and reinforces that CRISPR-Cas is not a single technology with a single risk-benefit profile but a family of platforms whose safety depends principally on delivery route and editing context (Raigani *et al.*, 2025).

For practice, the reasonable inference is that CRISPR-based assays are credible candidates for decentralised or resource-limited detection of well-characterised targets, particularly where their short time-to-result—a median of approximately 60 minutes in the MRSA review—offers advantages over centralised PCR. They are not yet supported by the kind of prospective, consecutively recruited, clinically embedded accuracy studies that would justify replacing an established reference method. For therapeutics, the evidence supports continued development under close safety surveillance, with a clear distinction maintained between the relatively well-characterised ex vivo route and the less predictable in vivo route.

5. Limitations

Several limitations constrain the interpretation of this review and are stated explicitly rather than minimised.

- The review was not prospectively registered and no published protocol exists, which is itself an AMSTAR-2 critical weakness and leaves the possibility of post hoc analytic decisions unaddressed.
- The unit of analysis is the systematic review, not the primary study. Pooling review-level summary estimates is a second-order synthesis with known limitations: it inherits every bias present in the constituent reviews, weights reviews rather than evidence, and cannot correct errors in the original pooling.
- Primary-study overlap between the three SARS-CoV-2 reviews could not be quantified because the reviews do not report their included-study lists in a form permitting matching; the corrected covered area was therefore not calculable. The non-overlapping primary analysis mitigates but does not eliminate this problem.
- Standard errors were derived one-sidedly from reported lower confidence limits because several upper limits were truncated at the 1.00 boundary. This is an approximation. It is likely to be conservative for near-

boundary estimates but has not been validated, and the specificity prediction interval in particular should be treated as indicative only.

- Point estimates reported as 1.00 were set to 0.999 before logit transformation. Results for specificity are consequently sensitive to this arbitrary choice, and a bivariate model fitted to primary 2×2 data would be preferable to the univariate approach used here.
- A bivariate or hierarchical summary ROC model, which is the methodologically preferred approach for diagnostic accuracy meta-analysis (Reitsma *et al.*, 2005) because it accounts for the correlation between sensitivity and specificity, could not be fitted because the included reviews do not report the covariance structure required.
- Publication bias was not assessed, as fewer than ten units of analysis were available. Given that near-perfect accuracy is more likely to be submitted and accepted than mediocre accuracy in a novel technology field, small-study and positive-result effects should be assumed present but unquantified.
- AMSTAR-2 items that could not be verified from the published record were marked 'unclear' rather than scored, which may render the appraisal conservative for some reviews and insufficiently critical for others.
- Searches were restricted to English-language publications, and a substantial portion of the CRISPR diagnostics primary literature originates from Chinese research groups; relevant non-English reviews may have been missed.
- The therapeutic component is a narrative synthesis drawing partly on regulatory announcements and conference reporting where peer-reviewed publication was not yet available. These sources are appropriately cited but carry a lower evidentiary standard than peer-reviewed trial reports, and the safety signal discussed remains under investigation at the time of writing.

6. Conclusion

CRISPR-Cas-based diagnostic assays demonstrate high pooled sensitivity (0.978, 95% CI 0.946–0.991) and specificity (0.995, 95% CI 0.967–0.999) across viral and bacterial targets. However, these estimates derive from systematic reviews of low or critically low methodological quality, synthesising predominantly laboratory-based accuracy studies with near-ceiling results, and are accompanied by prediction intervals wide enough that performance in a new pathogen or field setting cannot be confidently anticipated. The apparent replication across reviews is substantially attributable to shared primary studies rather than independent confirmation.

Therapeutic CRISPR-Cas genome editing has achieved genuine clinical milestones, including a first regulatory approval and durable biomarker responses from in vivo editing, but the trial evidence remains too sparse, too heterogeneous in endpoint and too short in follow-up to support quantitative pooling; a serious hepatotoxicity signal in a phase 3 in vivo programme during 2025 underscores that safety cannot be generalised across delivery platforms.

The priority for the field is not further meta-analysis of the existing diagnostic evidence base but prospective, consecutively recruited, clinically embedded accuracy studies reported to STARD standards (Bossuyt *et al.*, 2015), and registered non-redundant reviews thereof. For therapeutics, long-term follow-up for off-target and oncogenic outcomes, and randomised comparison against standard of care where ethically feasible, are the necessary next steps.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Acknowledgment

The authors thank the institutional library service for assistance with database access.

References

Abavisani, M., Karbas Foroushan, S., Khayami, R. and Sahebkar, A. (2024). Mycobacterium tuberculosis detection using CRISPR technology: An updated systematic review and meta-analysis. *Mol Diagn Ther*, 28(6): 777-790. doi: 10.1007/s40291-024-00741-x

Aromataris, E., Fernandez, R., Godfrey, C.M., Holly, C., Khalil, H. and Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc*, 13(3): 132-140.

Bossuyt, P.M., Reitsma, J.B., Bruns, D.E. *et al.* (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, 351: h5527.

Broughton, J.P., Deng, X., Yu, G. *et al.* (2020). CRISPR-Cas12-based detection of SARS-CoV-2. *Nat Biotechnol*, 38(7): 870-874.

Chen, J.S., Ma, E., Harrington, L.B. *et al.* (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, 360(6387): 436-439.

Deeks, J.J., Macaskill, P. and Irwig, L. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*, 58(9): 882-893.

DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Control Clin Trials*, 7(3): 177-188.

Farasati Far, B., Akbari, M., Habibi, M.A., Katavand, M. and Nasseri, S. (2025). CRISPR technology in disease management: An updated review of clinical translation and therapeutic potential. *Cell Prolif*, 58(11): e70099. doi: 10.1111/cpr.70099.

Frangoul, H., Altshuler, D., Cappellini, M.D. *et al.* (2021). CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *N Engl J Med*, 384(3): 252-260.

Gao, Y. and Chen, J. (2025). Fast but accurate: A systematic review and meta-analysis on diagnostic performance of MRSA detection in clinical samples by using CRISPR-based rapid molecular methods. *Front Microbiol*, 16: 1703247. doi: 10.3389/fmicb.2025.1703247.

Gillmore, J.D., Gane, E., Taubel, J. *et al.* (2021). CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis. *N Engl J Med*, 385(6): 493-502.

Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W. *et al.* (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, 356(6336): 438-442.

Grinstein, J.D. (2025). Reporter's notebook: Is Intellia's patient death the nail in the coffin for Cas9?. *Inside Precision Medicine*, 10 November.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414): 557-560.

Huang, P.Y., Yin, X., Huang, Y.T., Ye, Q.Q., Chen, S.Q., Cao, X.J., Xie, T.A. and Guo, X.G. (2022). Evaluation of CRISPR-based assays for rapid detection of SARS-CoV-2: A systematic review and meta-analysis. *Yonsei Med J*, 63(5): 480-489. doi: 10.3349/ymj.2022.63.5.480.

Innovative Genomics Institute. (2025). CRISPR clinical trials: A 2025 update. IGI, Berkeley, CA. Available from: https://innovativegenomics.org/news/crispr-clinical-trials-2025/

IntHout, J., Ioannidis, J.P.A., Rovers, M.M. and Goeman, J.J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7): e010247.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096): 816-821.

Longhurst, H.J., Lindsay, K., Petersen, R.S. *et al.* (2024). CRISPR-Cas9 in vivo gene editing of KLKB1 for hereditary angioedema. *N Engl J Med*, 390(5): 432-441.

Page, M.J., McKenzie, J.E., Bossuyt, P.M. *et al.* (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372: n71.

Pan, Z., Xu, L., Fan, Z. and Ren, F. (2025). CRISPR-Cas for hepatitis virus: A systematic review and meta-analysis of diagnostic test accuracy studies. *Front Microbiol,* 16: 1509890. doi: 10.3389/fmicb.2025.1509890.

Pieper, D., Antoine, S.L., Mathes, T., Neugebauer, E.A.M. and Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *J Clin Epidemiol*, 67(4): 368-375.

Raigani, M., Eftekhari, Z., Adeli, A. and Kazemi-Lomedasht, F. (2025). Advancing gene editing therapeutics: Clinical trials and innovative delivery systems across diverse diseases. *Mol Ther Nucleic Acids*, 36(3): 102666. doi: 10.1016/j.omtn.2025.102666.

Reitsma, J.B., Glas, A.S., Rutjes, A.W.S., Scholten, R.J.P.M., Bossuyt, P.M. and Zwinderman, A.H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*, 58(10): 982-990.

Shea, B.J., Reeves, B.C., Wells, G. *et al.* (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358: j4008.

Wang, S., Hu, J., Sui, C., He, G., Qu, Z., Chen, X., Wang, Y., Guo, D. and Liu, X. (2022). Accuracy of clustered regularly interspaced short palindromic repeats (CRISPR) to diagnose COVID-19, a meta-analysis. *Microb Pathog*, 165: 105498. doi: 10.1016/j.micpath.2022.105498

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E. *et al.* (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 155(8): 529-536.

Zhang, X., Ge, X., Shen, F., Qiao, J., Zhang, Y. and Li, H. (2022). Diagnostic efficiency of RPA/RAA integrated CRISPR-Cas technique for COVID-19: A systematic review and meta-analysis. *PLoS One*, 17(10): e0276728. doi:10.1371/journal.pone.0276728

Cite this article as: Wei-Chen Lin (2025). Pooled diagnostic accuracy of CRISPR-Cas nucleic acid detection assays: A second-order meta-analysis of published systematic reviews, with prediction intervals and leave-one-out sensitivity analysis. *Biomed Pharma Reviews.* 1(2), 25-38. doi: 10.62587/BMPR.1.2.2025.25-38.