The AI-in-RWE Transparency (AIRT) Checklist: Essential and Desirable Standards for AI-Enhanced Real-World Evidence

MSR195

Tushar Srivastava, MSc[1]; Radha Sharma, PhD[2]; Raju Gautam, PhD[1]; Madhusudan Kabra, MSc[3]

[1]ConnectHEOR, London, UK; [2]ConnectHEOR, Edmonton, AB, Canada; [3]MK Global Consulting, London, UK.

Email: tushar.srivastava@connectheor.com

BACKGROUND

Context: Artificial intelligence (AI) is becoming integral to analytics that transform real-world data (RWD) into real-world evidence (RWE) for regulatory, health technology assessment (HTA) decisions and to generate evidence for other objectives.[1-3] Several HTA agencies have released documents on AI integration for RWE in reimbursement submissions, including NICE, CDA, and IQWiG,[3] specifying its use for data analysis, especially producing structured data from unstructured RWD.[4,5] However, existing reporting checklists, originally designed for conventional statistical studies, seldom address the additional complexity introduced by machine-learning workflows.[6-10] In AI-enabled RWE, crucial elements such as model training, performance validation, explainability, and bias mitigation are often inconsistently reported or omitted, limiting reproducibility and weakening decision-maker confidence. The growing regulatory focus on trustworthy AI, exemplified by the European Union's AI Act, further reinforces the need for clear documentation of AI system development, validation, and governance, particularly in high-risk domains like healthcare.[11]

Objective: This study aimed to develop a structured, open-access checklist for transparent use of AI in RWE (AI-in-RWE Transparency [AIRT]) that distinguishes between items considered essential for credibility and those deemed desirable based on context.

METHODS

The AIRT checklist was developed through a multi-phase, mixed-method approach to ensure methodological rigor, transparency, and practical applicability.

1. **Evidence Mapping:** A structured desk search was conducted to identify key sources informing transparency standards for AI applications in RWE. This included:
   - Regulatory and HTA guidance (e.g., NICE, CDA, FDA)[4,6,7,12]
   - Methodological position papers and peer-reviewed publications [2,13,14]

2. **Stakeholder Engagement:** Findings from the evidence mapping informed the development of a semi-structured interview guide, which was used to conduct discussions with key opinion leaders from HTA agencies, life-science companies, academia, and patient groups. Interview transcripts were analyzed using an inductive qualitative approach to identify, consolidate, and prioritize items considered critical for AI transparency and credibility in RWE generation.

3. **Checklist Development:** A multidisciplinary team of experts in RWE, AI, health economics, and patient-centered outcomes developed the AIRT checklist by synthesizing evidence from literature mapping and stakeholder input to define key domains and transparency items. Each item was assessed for clarity, relevance, and feasibility to ensure practical use across a wide range of AI-RWE studies.

4. **Pilot Testing and Validation:** The refined checklist was pilot tested across AI-RWE case studies (n=3)[15-17] to evaluate clarity, completeness, and ease of use. Feedback from this testing assessed the checklist's alignment with regulatory expectations and its capacity to support transparent and credible reporting. The final checklist differentiates between essential items (must be reported) that are required for credibility and desirable (should be reported) items that can be adapted based on study objectives and data maturity.

RESULTS and DISCUSSION

- The emerging AIRT checklist captures reporting expectations across domains such as **study conceptualization, data source and collection, analyses and modeling, quality appraisal, interpretation and dissemination, and validation processes by human in loop component**. A companion digital tool is in development to guide users through each domain, generate structured transparency summaries, and highlight potential reporting gaps.
- Early pilot mapping of published studies indicates that **consistent use of AIRT could improve clarity and streamline both internal and external review and evaluation processes**. Consistent implementation of the **AIRT framework can bring methodological clarity, transparency, and reproducibility by explicitly defining when, where, and how AI was used**, detailing the validation procedures applied to assess its accuracy, and evaluating and standardizing the reporting of these processes.
- **Limitations:** The AIRT framework is currently conceptual and requires broader iteration and validation across diverse therapeutic areas and data types. Implementation consistency may vary due to differences in AI maturity, governance structures, and data accessibility across organizations.

	Essential Items	Desirable Items
Study conceptualization	Clearly state the intended use of AI (e.g., study design, statistical analysis, cohort identification, algorithmic prediction, report drafting, SAP development).	Define the primary objective and context of use (e.g., evidence generation, exploratory research). Describe study scope, endpoints, and rationale for AI integration over conventional methods.
Data Collection	Specify data source(s), and time frame (claims, EHRs, registries, etc.). Describe AI-enabled data identification and preprocessing steps (e.g., NLP extraction, image recognition, automated coding).	Provide detailed algorithm development for data curation and validation of the scientific rationale behind variable or feature selection.
Analyses & Modeling	Specify AI/ML model type(s) used (e.g., LLMs, random forests, neural networks) and the justification for selection. Provide model validation strategy (internal, external, cross-validation) and metrics used.	Document training datasets, preprocessing, and feature engineering methods. Discuss model applicability, interpretability, and performance boundaries.
Quality Appraisal	Identify potential sources of bias (data, algorithmic) and describe methods for detection and mitigation. Describe how biases were reported or addressed in study interpretation.	Report any disparities in model performance across subgroups (e.g., age, gender, ethnicity, disease severity).
Ethics	Indicate whether ethics or Institutional Review Board approval was required and obtained, and describe data privacy safeguards, consent management, and governance processes implemented for AI-driven analyses.	Disclose compliance with broader regulatory and data governance frameworks (e.g., GDPR, HIPAA), explaining how these standards were operationalized to ensure transparency, accountability, and responsible AI use.
Interpretation	Document performance validation and reproducibility testing (e.g., independent reruns, code verification). Include sensitivity analyses or uncertainty quantification methods relevant to AI outputs.	Present AI-derived results alongside conventional analyses, highlighting differences in outcomes or insights.
Dissemination	Ensure comprehensive reporting of AI methodology, including access to code, model parameters, and training data (where permissible). Declare limitations and uncertainties specific to AI model outputs.	Provide references or repositories for models or datasets used, when publicly available.
Human in Loop	Specify the role and timing of human expert input (during data curation, model training, validation, or interpretation). Describe mechanisms for human review of AI outputs, including override protocols or expert arbitration.	Report the qualifications or domain expertise of reviewers involved in human-AI interactions. Reflect on how human oversight improved transparency, reliability, or ethical balance.

AIRT — Advancing transparency and trust in AI-driven healthcare evidence. The AIRT checklist standardizes transparency, validation, and reporting practices for AI-enabled real-world evidence, ensuring methodological rigor, scientific credibility and reproducibility.

REFERENCES

1. CDA-AMC. Position Statement on the Use of Artificial Intelligence in the Generation and Reporting of Evidence. April 2025. Available at: https://www.cda-amc.ca/sites/default/files/MG%20Methods/Position_Statement_AI_Renumbered.pdf [Accessed: 21 Sept 2025]
2. Fleurence, et al. Value Health. 2025; 28(2): 175-183.
3. Ting, et al. Value Health. 2025; 28(6): S295-S296.
4. NICE. Statement of intent for artificial intelligence. 2024. Available at: http://www.nice.org.uk/corporate/ecd12/resources/nicestatement-of-intent-for-artificial-intelligence-ai-pdf-4046427062394l. [Accessed: 6 Aug 2025]
5. CDA-AMC. Transforming Real-World Evidence with AI: From Data to Decisions. 15 Oct 2025. Available at: www.cda-amc.ca/events/transforming-real-world-evidence-ai-data-decisions [Accessed 25 Oct 2025]
6. CDA-AMC. Checklist for Reporting Real-World Evidence (MG0020). May 31 2023. Available at: https://www.cda-amc.ca/sites/default/files/RWE/MG0020/MG0020-Final-Checklist-e.pdf (accessed Aug 2025)
7. CDA-AMC. Guidance for Reporting Real-World Evidence (MG0020-RWE Guidance Report). May 31 2023. Available at: https://www.cda-amc.ca/sites/default/files/RWE/MG0020/MG0020-RWE-Guidance-Report-Secured.pdf [Accessed Aug 2025).
8. Wang, et al. BMJ. 2021; 372.
9. Fleurence, et al. Value Health. 2024; 27(6): 692-701.
10. Berger, et al. Value Health. 2017; 20(8): 1003-1008.
11. European Parliament. (2024) [Accessed 31 July 2024]. https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf.
12. U.S. Food & Drug Administration. *Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products: Guidance for Industry and Other Interested Parties* (Draft). Jan 2025.
13. Naylor, et al. Clinical Translational Sci. 2025; 18(4): e70206.
14. Srivastava, et al. Advances Therapy. 2025: 1-28.
15. Maddali, et al. The Lancet Respiratory Med. 2022; 10(4): 367-377.
16. Sau, et al. The Lancet Digital Health. 2024; 6(11): e791-e802.
17. Flores, et al. Circulation Res. 2021; 128(12): 1833-1850.

ACKNOWLEDGEMENTS

The authors acknowledge the support of KOLs who participated in discussions and reviews that helped shape the AIRT framework and ensure its practical relevance.

FINANCIAL DISCLOSURE

Authors are employees or external collaborators of ConnectHEOR Limited. No external funding was received for this research, and the authors declare no conflicts of interest.

Poster presented at ISPOR in Glasgow, Scotland, 9-12 November 2025