

Can readers understand findings from systematic reviews almost as well from artificial intelligence-assisted visual abstracts as from those published by The BMJ?

Submission date 26/08/2026	Recruitment status Not yet recruiting	☒ Prospectively registered ☐ Protocol
Registration date 09/09/2026	Overall study status Ongoing	☐ Statistical analysis plan ☐ Results
Last Edited 09/09/2026	Condition category Other	☐ Individual participant data ☒ Record updated in last year

Plain English summary of protocol

Background and study aims

Systematic reviews bring together the results of many studies to answer a healthcare question. They are useful but are often long and technical. Visual abstracts use short text and simple graphics to show the main points more quickly and clearly, but making them takes time and design skill.

Artificial intelligence (AI) could help, but it can also make mistakes, leave out how certain or uncertain the evidence is, or present findings in a misleading way. Our study tests a structured way of working that aims to reduce these risks. AI will produce a first draft from a fixed set of written instructions. Two researchers experienced in systematic reviews will check each draft against the original review and correct any problems.

Our main question is whether readers understand the information almost as well from the AI-assisted visual abstracts as from visual abstracts published by The BMJ.

Who can participate?

We will recruit 656 adults aged 18 years or older, of any sex, who have relevant professional, academic, policy-related, or actively engaged healthcare or research experience. This may include clinicians, health researchers, public health professionals, systematic review authors or users, guideline developers, policymakers, healthcare decision makers, and eligible patient, caregiver, or peer-research partners actively involved in healthcare or research initiatives. Participants must be able to read and understand English, have internet access and a suitable device, and provide electronic informed consent. People who helped select, develop, anonymise, pilot, or pretest the visual abstracts used in the trial cannot participate.

What does the study involve?

A computer will randomly decide whether each participant sees three AI-assisted visual abstracts or three visual abstracts already published by The BMJ. Both sets will cover the same three systematic reviews.

After each visual abstract, participants will answer six questions about the information shown

and rate its clarity and trustworthiness. Each participant will complete these assessments online in a single session.

Separately, the research team will assess the reading level of the text. Independent experts will compare the final visual abstracts in both groups with their source reviews and judge whether they could mislead readers. They will also compare each first uncorrected AI-generated draft with its source review and record the types, number, and locations of any errors.

What are the possible benefits and risks of participating?

Participants should not expect a direct personal or clinical benefit. Their responses may help researchers improve how findings from systematic reviews are communicated. The study does not involve medical treatment or physical procedures. The main burdens are the time and concentration needed to view the materials and answer the questions, and possible tiredness or frustration while completing the online assessments.

Where is the study run from?

The study is managed by the Centre for Health Research Methodology at the University of Galway, Ireland. Participants will be recruited internationally through Prolific and will complete all procedures remotely from their own locations.

When is the study starting and how long is it expected to run for?

November 2026 to March 2027.

Who is funding the study?

The study is funded by the University of Galway (Ireland).

Who is the main contact?

Xuenan Pang

x.pang1@universityofgalway.ie

Contact information

Type(s)

Principal investigator, Scientific, Public

Contact name

Ms Xuenan Pang

ORCID ID

<https://orcid.org/0000-0002-5577-237X>

Contact details

Centre for Health Research Methodology, School of Nursing, Midwifery and Evidence Science,
University of Galway

Galway

Ireland

H91 TK33

+353 91494492

x.pang1@universityofgalway.ie

Additional identifiers

Study information

Scientific Title

Comparison of AI-assisted visual abstracts and published BMJ visual abstracts for systematic reviews: protocol for a non-inferiority randomised trial

Acronym

HIET-3

Study objectives

HIET-3 aims to determine whether AI-assisted visual abstracts, developed through a locked LLM-assisted workflow and human verification, can communicate evidence from systematic reviews to international healthcare interest holders as effectively as existing published BMJ visual abstracts.

The primary objective is to test non-inferiority in evidence comprehension of shared core information. Secondary objectives are to compare the two arms descriptively on graphic clarity and suitability, perceived trustworthiness, textual readability, and communicative safety, and to describe AI-generation accuracy as a process outcome. The trial will also explore participants' perceptions of how each visual abstract was developed and how important source attribution is to them.

Ethics approval required

Ethics approval required

Ethics approval(s)

Approved 11/10/2024, University of Galway Research Ethics Committee (University Road, Galway, H91 TK33, Ireland; +353 (0)91 524411; ethics@universityofgalway.ie), ref: 2023.05.011 Amend 2410

Primary study design

Interventional

Allocation

Randomized controlled trial

Masking

Blinded (masking used)

Control

Active

Assignment

Parallel

Purpose

Health information communication and evidence dissemination

Study type(s)

Health condition(s) or problem(s) studied

Health information communication and comprehension of findings from systematic reviews presented in visual abstracts.

Interventions

This study is a randomised, parallel-group, two-arm, non-inferiority trial among international healthcare interest holders. It compares visual abstracts produced through a prespecified, checklist-informed, human-verified AI-assisted workflow with the selected visual abstracts published by The BMJ.

Participants will be randomised in a 1:1 allocation ratio to one of two groups:

Arm 1 — AI-assisted visual abstracts

Arm 2 — BMJ comparator visual abstracts

Sequence generation

After electronic informed consent and eligibility confirmation, participants will be assigned using QuestionPro's survey Block Randomiser tool, configured to display one of two arm-specific survey blocks. Despite the name of the platform feature, the trial will use computer-generated simple randomisation rather than permuted-block randomisation. Each participant will be assigned independently with equal probability to either the AI-assisted visual abstract arm or the BMJ comparator visual abstract arm. The displayed block (study arm) will be recorded in the platform data export.

Allocation concealment mechanism

QuestionPro will be configured to automatically direct eligible participants to one of the two study arms after consent and eligibility confirmation. Allocation will not be based on participant-identifying information, demographic characteristics, or responses to baseline questions. The research team will not have access to the allocation sequence at the point of enrolment.

AI-assisted visual abstracts

The intervention materials will consist of three LLM-assisted, human-verified visual abstracts developed from the same three source systematic reviews used for the comparator materials. The intervention development process will include four stages. These stages are presented in sequence for clarity, but development may involve iteration between generation, verification, and feedback before the final materials are locked.

1. Prompt development and iteration. The intervention will be generated using a locked prompt package with two sequential components. The first component will focus on the textual content, including the study question, population, intervention, comparator, outcomes, main findings, uncertainty, and take-home message. The second component will focus on the layout and visual presentation. It will specify a common visual structure across intervention materials while allowing limited article-specific adaptation, such as illustrations that reflect the topic or population of the source review. To align the intervention with published visual abstract standards, the prompt package will be informed by the Reporting Infographics and Visual Abstracts of Comparative studies (RIVA-C) checklist and guide and the Ibrahim Visual Abstract Primer. These sources provide general design principles and recommended content elements rather than the specific comparator materials. Prompt development and calibration will be conducted using systematic reviews that will not be included in the final trial materials. The prompt package will be refined to improve clarity, completeness, consistency, and faithfulness to the source article. Existing BMJ visual abstracts linked to the selected source reviews will not be provided to the LLM during calibration, generation, or refinement and will not be used to guide revisions to the AI-assisted materials.

2. LLM generation. After calibration, the prompt package will be locked and applied without substantive modification to the three selected source reviews. For each review, the content prompt will first be used to generate the display text. The layout prompt will then be used to

generate the visual abstract plan and the first complete rendered draft based on the approved content. The exact model and complete tool chain, including each tool's function and position in the workflow, will be prespecified and locked before any selected source review is processed. The same model and tools will be applied consistently across all three source reviews, with no substitution after generation begins. For each generation, the research team will record the LLM name, model name and version and any documented training data or knowledge cutoff, date of use, interface or platform, prompt package version, full prompt text, input materials, output files, and any available generation settings, such as temperature, seed, image size or other configurable parameters. The research team will also record the rendering or design software used to produce the first complete rendered draft, software version (where available), output file format, and any separate image-generation, image-editing, or design-assistance tools that materially contributed to that draft.

3. Human verification and correction. Two experts in systematic review methodology will independently review each first complete rendered LLM-generated visual abstract draft against the source article. They will check whether the draft accurately reflects the study design, population, intervention, comparator, outcomes, main findings, uncertainty, and important limitations or safety-related information. Required corrections will be documented in a development correction log before revision. This log will record each verification and correction round, including the date, verifier name, issues identified, requested corrections, whether the corrections are made by regeneration or manual editing, software or tools used for revision, conflict resolution, revised output files, and the reason for stopping refinement. This log will document the intervention development process and will not be available to outcome assessors. The experts who verify and correct the intervention materials will not serve as independent outcome raters for safety or AI-generation accuracy. A maximum of three refinement rounds will be conducted for each LLM-assisted visual abstract. A refinement round is defined as expert review of the current draft against the source article, followed by correction or regeneration where required. If regeneration is needed, the locked prompt and source materials will be reused, and the revised output will undergo further human verification. Refinement will stop before the maximum number of rounds if all of the following criteria are met: (1) no unresolved content-inaccuracy errors remain, including incorrect, partially incorrect, hallucinated, or unsupported content; (2) no unresolved communicative safety concerns remain, including risks of misinterpretation, inappropriate recommendations, omission of important uncertainty, or inconsistency with the source article; and (3) further language refinement produces less than a 5% change in the Flesch–Kincaid Grade Level compared with the previous round. The textual readability criterion will be applied only after content accuracy and communicative safety have been addressed. If a content-inaccuracy or communicative-safety concern remains after the third round, only the correction needed to resolve it will be made and independently rechecked before the material is locked; any additional round will be documented.

4. Target-user feedback. After expert verification and correction, the revised visual abstracts will be reviewed by a small target-user and patient and public involvement (PPI) feedback panel. We will aim to include approximately five contributors, including target users and two PPI contributors. Target users will be individuals who resemble the intended readers of the journal comparators. PPI contributors will provide a public or non-specialist perspective, helping to identify unexplained terminology and other barriers for a lay audience. Based on expert verification and target-user feedback, revisions will be limited to prespecified corrections or refinements that address factual accuracy, unsupported claims, communicative safety, clarity, readability, usability, and display errors. Revisions will not be made to mimic the style, emphasis, structure, or visual choices of the journal comparators. The panel will not assess study outcomes or participate in the main trial. Any changes made after feedback will be documented and, if they affect text, numerical content, interpretation, or visual emphasis, rechecked against the source article by the two methodology experts before the final intervention materials are locked and participant recruitment begins.

BMJ comparator visual abstracts

The comparator materials will consist of three existing BMJ visual abstracts linked to systematic reviews of intervention studies. These materials will form the BMJ comparator set. Their production histories cannot be directly observed and may include editorial, design, and software-assisted steps.

Eligible comparator visual abstracts must be listed in The BMJ's official visual abstracts archive [<https://www.bmj.com/content/bmj-visual-abstracts>], be linked to a systematic review of human healthcare intervention studies, and have been published before 19 October 2023. The BMJ was selected because it publishes visual abstracts linked to clinically relevant systematic reviews and has an international health-professional readership. This cutoff was chosen because OpenAI announced on that date that DALL·E 3 had become available in ChatGPT Plus and Enterprise. Any candidate visual abstract for which a publicly available declaration identifies AI use in its production or design will be excluded. This criterion was used to reduce the likelihood that ChatGPT-integrated image generation contributed to the comparator materials, although it cannot rule out all forms of AI or software-assisted design.

Two reviewers will independently examine all archive entries and assess them against these criteria. They will compare their records and resolve discrepancies in identified entries or eligibility decisions through discussion, consulting a third reviewer if required. Reasons for exclusion will be recorded. If more than three eligible BMJ visual abstracts are identified, the final set will be selected using a prespecified maximum-variation procedure based on two characteristics of the source reviews: broad healthcare topic and review complexity. Broad healthcare topic will be defined by the principal clinical area or health condition addressed by the review. A review will be classified as complex if it meets at least two of the criteria proposed by Mahtani et al. The research team will first identify sets of three visual abstracts containing the greatest possible number of distinct healthcare topics. If more than one set meets this criterion, preference will be given to sets containing at least one complex and one non-complex review. If multiple sets remain eligible, one will be selected using a computer-generated random draw. The candidate classifications and random-selection record will be retained. The selected comparators will remain concealed from the prompt-development and intervention-refinement team until the AI-assisted materials are locked, with comparator selection undertaken by other team members.

All participant procedures will be completed online in a single study session. On the day of participation, potential participants will be recruited through Prolific and directed to QuestionPro. After reviewing the study information and providing electronic informed consent, potential participants will complete the eligibility assessment in QuestionPro. Eligible participants will then complete baseline and demographic questions and will then be randomly assigned by the survey platform in a 1:1 ratio to either the AI-assisted visual abstract arm or the BMJ comparator arm. Participants in both arms will view visual abstracts based on the same three source systematic reviews. After viewing each visual abstract, participants will complete the visual-abstract-specific questionnaire: evidence comprehension, graphic clarity/suitability, and perceived trustworthiness. After completing the assessments for all three visual abstracts, participants will answer two sets of additional exploratory questions. Back-navigation to the main outcome assessments will be disabled at the start of the exploratory block. Participants will be instructed to answer based only on the visual abstract presented, not to search for the source reviews during the session, and will not be provided with the source article, citation, or external links.

All data will be collected remotely. The estimated total participation time is 30-60 minutes, including informed consent, eligibility confirmation, answering demographic questions, viewing

three visual abstracts, and completing all outcome assessments. This estimate, and the participant payment based on it, will be confirmed from the pilot completion times before recruitment begins.

Intervention Type

Other

Primary outcome(s)

1. Evidence comprehension measured using six source-review-specific multiple-choice questions per visual abstract, identical across study arms, with item-level responses scored as correct or incorrect at three timepoints during the single online session, immediately while viewing each visual abstract

Key secondary outcome(s)

1. Graphic clarity and suitability measured using five adapted items informed by the Suitability Assessment of Materials, each rated on a 5-point Likert scale; the visual-abstract-level mean is the main measure, with higher scores indicating greater perceived graphic clarity and suitability at three timepoints during the single online session, after the comprehension questions for each visual abstract

2. Perceived trustworthiness measured using five adapted trustworthiness items rated on a 5-point Likert scale; the visual-abstract-level mean is the main measure, with higher scores indicating greater perceived trustworthiness at three timepoints during the single online session, after the graphic clarity and suitability items for each visual abstract

3. Textual readability measured using seven indices applied to text extracted from each final anonymised visual abstract: Flesch Reading Ease, Flesch–Kincaid Grade Level, Gunning Fog Index, Automated Readability Index, Coleman–Liau Index, SMOG Index, and Linsear Write Formula; Flesch–Kincaid Grade Level is the main measure at one timepoint after anonymisation and final material locking

4. Safety measured using four core criteria assessed as met or not met by two independent systematic review methodology experts comparing each final anonymised visual abstract in both study arms with its source article; the main measure is the percentage of criteria met, with concerns classified as minor, moderate, major, or critical at one timepoint after final material locking

5. AI-generation accuracy measured using an adapted error schema applied by two independent systematic review methodology experts to each first complete rendered AI-generated visual abstract draft, produced before human verification or correction, and its source article; error categories, frequencies, and locations are recorded; AI-assisted arm only at one timepoint after the safety ratings have been completed and locked if the same experts assess both outcomes

6. Additional exploratory questions measured using two sets of exploratory questions: (1) perceived development source of the set of visual abstracts, with categorical response options; and (2) importance of source attribution, rated on a 5-point scale at one timepoint after all three visual-abstract assessment blocks

Completion date

31/03/2027

Eligibility

Key inclusion criteria

1. An international healthcare interest holder, defined for this trial as an individual with professional, academic, or policy-related experience in healthcare, clinical practice, public health, health research, evidence synthesis, guideline development, health technology assessment, journal publishing, or healthcare policy. Eligible groups may include clinicians, health researchers, public health professionals, systematic review authors or users, guideline developers, policymakers, healthcare decision makers, patient/caregiver or peer-research partners actively involved in healthcare or research initiatives, and professionals involved in evidence-informed healthcare practice or policy, provided they meet the experience-based definition above
2. Aged 18 years or older
3. Of any sex
4. Able to read and understand English sufficiently to engage with the visual abstract materials and complete the study assessments. Participants must report English as their native or first language or select Fluent, Advanced, or Mastery on the adapted Common European Framework of Reference for Languages-based screening item
5. Access to the internet and to a computer, tablet, smartphone, or other device capable of displaying the visual abstract materials and completing the online study procedures
6. Able and willing to provide electronic informed consent before participation

Healthy volunteers allowed

No

Age group

Mixed

Lower age limit

18 Years

Upper age limit

100 Years

Sex

All

Total final enrolment

0

Key exclusion criteria

Previous participation in the selection, development, anonymisation, piloting, or pretesting of the visual abstract materials used in the trial

Date of first enrolment

01/11/2026

Date of final enrolment

22/12/2026

Locations

Countries of recruitment

United Kingdom

England

Australia

Brazil

Canada

Germany

Ireland

Italy

Mexico

Poland

Portugal

South Africa

Spain

United States of America

Study participating centre

Online recruitment through Prolific

Fully remote online study. International participants will be recruited through Prolific and will complete all consent and study procedures remotely from their own location. There is no physical UK recruitment centre.

NA – online study

England

NA

Study participating centre

International online recruitment through Prolific

This is a fully remote international study open to eligible participants in all countries supported by Prolific. The 13 countries listed under Countries of recruitment are those most likely to contribute participants, and do not represent geographical eligibility restrictions. Participants will complete all consent and study procedures remotely from their own locations. There is no physical recruitment centre outside the UK.

NA – online study

Ireland
NA

Sponsor information

Organisation

Ollscoil na Gaillimhe – University of Galway

ROR

<https://ror.org/03bea9k73>

Funder(s)

Funder type

Funder Name

University of Galway

Alternative Name(s)

Coláiste na hOllscoile, Gaillimh, Ollscoil na hÉireann Gaillimh, Queen's College, Galway, University College, Galway, NUI Galway, National University of Ireland, Galway, National University of Ireland Galway, Ollscoil na Gaillimhe, National University of Ireland, Galway/NUI Galway, OÉ Gaillimh, NUI Galway

Funding Body Type

Government organisation

Funding Body Subtype

Universities (academic only)

Location

Ireland

Results and Publications

Individual participant data (IPD) sharing plan

De-identified data may be accessed by authorised members of the research team for data cleaning and analysis. Study data will be downloaded from the online platforms and stored in password-protected files on secure institutional storage (i.e., University of Galway's OneDrive). Data will be handled in accordance with the applicable requirements of the General Data Protection Regulation and the University of Galway data protection policies. De-identified data will be made available from the corresponding author upon reasonable request.

IPD sharing plan summary

Available on request, Stored in non-publicly available repository

Study outputs

Output type	Details	Date created	Date added	Peer reviewed?	Patient-facing?
Participant information sheet			28/08/2026	No	Yes