ISRCTN28623546 https://doi.org/10.1186/ISRCTN28623546

Helping people understand health research: AI podcasts versus human-produced podcasts

Submission date
15/08/2025

Recruitment status
No longer recruiting

Registration date
02/09/2025

Overall study status
Completed

Last Edited
16/11/2026

Condition category
Other

[_] Prospectively registered
[X] Protocol
[_] Statistical analysis plan
[_] Results
[_] Individual participant data
[X] Record updated in last year

Plain English summary of protocol
Background and study aims
Effective communication of health information allows for informed decision-making. Plain Language Summaries (PLSs) are short, easy-to-read versions of research reviews that explain what the evidence shows in everyday language. Thanks to advances in artificial intelligence (AI), there are now new and possibly more engaging ways to share this information, like AI-created podcasts. These podcasts may help people understand health research better, but how effective they really are is unknown. Generating evidence about their effectiveness is important to ensure that this new method of communication genuinely improves public understanding. This study aims to assess whether AI-assisted podcasts are as effective as traditional PLSs. This is assessed through comprehension, listenability, quality of information, perceived trustworthiness, and safety. Results relate to the AI-assisted audio delivery approach rather than AI alone.

Who can participate?
Healthy volunteers aged 18 years and over from anywhere in the world can enrol online at the University of Galway (Ireland)

What does the study involve?
Participants will be recruited via Prolific, an audience recruitment platform. People who decide to join will randomly be placed into two groups. Both groups will be asked to listen to podcasts talking about health. Participants don't need to know much about health topics. Both the podcast and the written summary will be easy to understand and explain the health topics clearly.
To find out how well people learned from the summaries, they will be asked to answer multiple-choice questions about their understanding of the material. The results of the two groups will be compared to see if there are any differences.

What are the possible risks and benefits of taking part?
There are no direct benefits from taking part in the study, but participants' answers will help us learn if artificial intelligence can be used to create health information that is easy for people to understand. They also might have some fun!

There are no major risks in taking part in this study. However, sometimes, people might feel uncomfortable if the information they are presented with doesn't match their own experiences or beliefs. To make everyone feel comfortable, there will be given an information leaflet explaining the study so they know what to expect and can choose if they want to join. They can also decide to leave the study anytime they want, with no consequences at all.

Where is the study run from?
University of Galway (Ireland)

When is the study starting and how long is it expected to run for
October 2024 to April 2026

Who is funding the study?
College of Medicine, Nursing and Health Sciences, University of Galway (Ireland)

Who is the main contact?
Isabel O'Byrne, i.obyrne1@universityofgalway.ie

Contact information

Type(s)
Public, Scientific, Principal investigator

Contact name
Ms Isabel O'Byrne

Contact details
University of Galway
Galway
Ireland
H91 TK33
-
i.obyrne1@universityofgalway.ie

Additional identifiers

Protocol serial number
Health Research Board (HRB) and Health Service Executive (HSE) Grant INFO-2021-001

Study information

Scientific Title
Comparison of AI-assisted and human-produced podcasts derived from Cochrane PLSs: protocol for a randomised noninferiority trial (HIET2)

Acronym
HIET2

Study objectives

Current study objectives as of 04/02/2026:
Cochrane Plain Language Summaries (PLSs) aim to make evidence from systematic reviews accessible to the public, but written formats may not meet the needs of all users, particularly those with reading difficulties or a preference for auditory content. AI-assisted podcasts offer a scalable, low-cost alternative, but their effectiveness compared to human-written PLSs is not yet known. This study aims to evaluate whether AI-assisted podcasts are a non-inferior way to communicate Cochrane evidence, with the goal of improving accessibility and user engagement. This is a pragmatic comparison of an AI-assisted audio modality with human-generated podcasts; results pertain to the overall delivery approach (modality plus AI assistance), not AI in isolation.

Previous study objectives:
Cochrane Plain Language Summaries (PLSs) aim to make evidence from systematic reviews accessible to the public, but written formats may not meet the needs of all users, particularly those with reading difficulties or a preference for auditory content. AI-assisted podcasts offer a scalable, low-cost alternative, but their effectiveness compared to human-written PLSs is not yet known. This study aims to evaluate whether AI-assisted podcasts are a non-inferior way to communicate Cochrane evidence, with the goal of improving accessibility and user engagement. This is a pragmatic comparison of an AI-assisted audio modality with human-generated text; results pertain to the overall delivery approach (modality plus AI assistance), not AI in isolation.

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 91 524411; ethics@universityofgalway.ie), ref: 2023.05.011

Primary study design
Interventional

Study design
Randomized parallel-group two-armed non-inferiority trial

Study type(s)
Other

Health condition(s) or problem(s) studied
Health information

Interventions
Current interventions as of 04/02/2026:
This study is a randomised, parallel-group, two-arm, non-inferiority trial comparing the effectiveness of AI-assisted podcasts with human-generated podcasts.

Arm 1 - AI-assisted podcasts
Participants in the intervention group will receive three AI-assisted podcasts, created using a human-in-the-loop process that combines large language model scripting with high-quality text-to-speech voices. The podcasts are based solely on Cochrane Plain Language Summaries and cover diverse and relevant health topics of varying complexity. Each script undergoes expert review and iterative refinement to ensure clarity, accuracy, and accessibility. The final content is reviewed by a topic expert and a patient/public partner, with all changes documented to

maintain quality and transparency. Each podcast 6-8 minutes; scripts derived solely from Cochrane PLS text; human-in-the-loop expert review, topic-expert check, and PPI review; process documented.

Arm 2 - Human-generated podcasts
Participants in the control group will receive three human-produced podcasts based on the same three Cochrane Plain Language Summaries as the intervention group. Scripts will be drafted by an experienced science communicator using only the PLS and the same structured brief. Human narrators will record the episodes. Editorial revisions will follow the identical multireviewer workflow and stopping rules used in the AI-assisted arm. The same PICO/term clarity checklist used during the verification pass will be applied, but without AI assistance. Edits will be documented in the same categories (accuracy, definitions/clarity, style) to maintain consistency

Randomisation
Participants will be randomly assigned in a 1:1 ratio to either the control group (receiving human-generated podcasts) or the intervention group (receiving AI-assisted podcasts) using the Block Randomiser feature in QuestionPro. This process is fully automated, with no identifying information influencing group allocation. Each group will receive the same three podcasts, respectively. The randomisation logic in the Block Flow tab ensures equal distribution between groups and maintains consistency in the information presented to participants across both conditions.

Outcomes
The study's outcomes are guided by the QUEST framework (Quality, Expression, Safety, and Trust), which evaluates healthcare AI tools across dimensions such as understanding, information quality, safety, and trust.

Previous interventions:
This study is a randomised, parallel-group, two-arm, non-inferiority trial comparing the effectiveness of AI-assisted podcasts with human-generated PLSs.

Arm 1 - AI-assisted podcasts
Participants in the intervention group will receive three AI-assisted podcasts, created using a human-in-the-loop process that combines large language model scripting with high-quality text-to-speech voices. The podcasts are based solely on Cochrane Plain Language Summaries and cover diverse and relevant health topics of varying complexity. Each script undergoes expert review and iterative refinement to ensure clarity, accuracy, and accessibility. The final content is reviewed by a topic expert and a patient/public partner, with all changes documented to maintain quality and transparency. Each podcast ≤8 minutes; scripts derived solely from Cochrane PLS text; human-in-the-loop expert review, topic-expert check, and PPI review; process documented.

Arm 2 - Human-generated written summaries
Participants in the comparator group will receive the published Cochrane Plain Language Summaries (PLSs) for the same three reviews. These summaries are written by systematic review experts following Cochrane's standardised guidelines to ensure clarity, consistency, and accessibility. They avoid jargon and use plain language, structured under standard headings such as key messages, findings, confidence in the evidence, and limitations. They undergo standard peer review before publication and represent the best current practice in creating health information summaries.

Randomisation
Participants will be randomly assigned in a 1:1 ratio to either the control group (receiving human-generated summaries) or the intervention group (receiving AI-assisted podcasts) using the Block Randomiser feature in QuestionPro. This process is fully automated, with no identifying information influencing group allocation. Each group will receive the same three summaries or podcasts, respectively. The randomisation logic in the Block Flow tab ensures equal distribution between groups and maintains consistency in the information presented to participants across both conditions.

Outcomes
The study's outcomes are guided by the QUEST framework (Quality, Understanding, Expression, Safety, and Trust Understanding), which evaluates healthcare AI tools across dimensions such as understanding, information quality, safety, and trust.

Intervention Type
Other

Primary outcome(s)
Current primary outcomes as of 04/02/2026:
Comprehension of participants will be measured using a 10-question multiple-choice quiz after each podcast, assessing their understanding of the content. Questions cover topic knowledge, review methods, main results, evidence quality, and currency. The primary outcome is total comprehension score (out of 10).

Non-inferiority margin:
A difference of no more than 1 point (10 percentage points) between groups will define non-inferiority. For example, if the human-podcast scores an average of 8/10, the AI-assisted podcast will be considered non-inferior if its lower 95% confidence interval exceeds 7/10.

Previous primary outcomes:
Comprehension of participants will be measured using a 10-question multiple-choice quiz after each summary or podcast, assessing their understanding of the content. Questions cover topic knowledge, review methods, main results, evidence quality, and currency. The primary outcome is total comprehension score (out of 10).

Non-inferiority margin:
A difference of no more than 1 point (10 percentage points) between groups will define non-inferiority. For example, if the human-generated PLS scores an average of 8/10, the AI-assisted podcast will be considered non-inferior if its lower 95% confidence interval exceeds 7/10.

Key secondary outcome(s)
Current secondary outcomes as of 04/02/2026:
1. Format accessibility will be evaluated using both objective and subjective measures:
Objective readability will be measured using the Flesch-Kincaid Grade Level, applied to human-generated podcast transcripts (control) and AI-assisted podcast transcripts (interventions). This metric estimates the U.S. school grade level required to understand the content. A non-inferiority margin of 1 grade level will be applied.
Items assessing tone, clarity, structure/pacing, and engagement will be measured using a 5-item, 5-point Likert questionnaire. Mean scores will be compared using a 0.5-point non-inferiority margin, with a focus on format-appropriate accessibility rather than identical measures across formats.

2. Quality of Information will be measured by two review experts who will independently assess for errors (e.g., incorrect or missing info) by comparing podcasts to the original Cochrane review. A non-inferiority margin of 0.3 points on the 3-point scale will be applied, with AI-assisted podcasts considered non-inferior if their mean quality score is no more than 0.3 points lower than that of the human-produced podcasts.
3. Safety will be measured by expert raters who will check for risks like misinterpretation, bias, or hallucinations. Core safety criteria will be assessed across both formats, with a 10% non-inferiority margin. AI-specific safety issues will be reported descriptively.
4. Perceived trustworthiness will be self-reported by participants who will rate trust using a 5-item Likert scale (e.g., "I trust the information"). A mean score with a 0.5-point non-inferiority margin will be used to compare groups.

Previous secondary outcomes:
1. Format accessibility will be evaluated using both objective and subjective measures:
Objective readability will be measured using the Flesch-Kincaid Grade Level, applied to written PLS (control) and AI-assisted podcast transcripts (interventions). This metric estimates the U.S. school grade level required to understand the content. A non-inferiority margin of 1 grade level will be applied.
Items assessing tone, clarity, structure/pacing, and engagement will be measured using a 5-item, 5-point Likert questionnaire. Mean scores will be compared using a 0.5-point non-inferiority margin, with a focus on format-appropriate accessibility rather than identical measures across formats.
2. Quality of Information will be measured by two review experts who will independently assess for errors (e.g., incorrect or missing info) by comparing summaries and podcasts to the original Cochrane review. A 10% non-inferiority margin will be used based on an assumed 80% baseline accuracy for human summaries.
3. Safety will be measured by expert raters who will check for risks like misinterpretation, bias, or hallucinations. Core safety criteria will be assessed across both formats, with a 10% non-inferiority margin. AI-specific safety issues will be reported descriptively.
4. Perceived trustworthiness will be self-reported by participants who will rate trust using a 5-item Likert scale (e.g., "I trust the information"). A mean score with a 0.5-point non-inferiority margin will be used to compare groups.

Completion date
30/04/2026

Eligibility

Key inclusion criteria
1. Age: Must be 18 years or older.
2. Language Proficiency: Must be proficient in English.
3. Participants will self-assess their English proficiency on a scale from 1 (not comfortable) to 10 (very comfortable). Only those who rate themselves as 7 or higher will be eligible.
4. Internet Access: Must have access to the internet.
5. Device Access: Must have a device (e.g., computer, tablet, or smartphone) to access study materials.
6. Consent: Must provide informed consent before starting the study.

Participant type(s)
Healthy volunteer

Healthy volunteers allowed
Yes

Age group
Mixed

Lower age limit
18 Years

Upper age limit
100 Years

Sex
All

Total final enrolment
0

Key exclusion criteria
Current exclusion criteria as of 04/02/2026:
1. Survey Completion Issues: Participants will be excluded if they are unable to complete the online survey.
2. Fast responses: Responses will be excluded if the participant completes the study in under 25 minutes.
3. Straight-lining: Responses will be excluded if there is evidence of straight-line answering, defined as selecting the same response option for ≥80% of Likert-scale items in a single assessment block (excluding comprehension questions).

Previous exclusion criteria:
1. Survey Completion Issues: Participants will be excluded if they are unable to complete the online survey.
2. Fast responses: Responses will be excluded if the participant completes the study in under 15 minutes.
3. Straight-lining: Responses will be excluded if there is evidence of straight-line answering, defined as selecting the same response option for ≥80% of Likert-scale items in a single assessment block (excluding comprehension questions).

Date of first enrolment
01/06/2025

Date of final enrolment
01/04/2026

Locations

Countries of recruitment
Ireland

Study participating centre

Participants will be recruited via an audience recruitment platform
School of Medicine
University of Galway,
University Road
Galway
Ireland
H91 TK33

Sponsor information

Organisation
Ollscoil na Gaillimhe – University of Galway

ROR
https://ror.org/03bea9k73

Funder(s)

Funder type
Not defined

Funder Name
College of Medicine, Nursing and Health Sciences, University of Galway

Alternative Name(s)
College of Medicine, Nursing and Health Sciences, National University of Ireland, Galway, College of Medicine Nursing & Health Sciences, NUI Galway - College of Medicine, Nursing and Health Sciences, College of Medicine, Nursing & Health Sciences - NUI Galway

Funding Body Type
Government organisation

Funding Body Subtype
Universities (academic only)

Location
Ireland

Funder Name
Health Research Board

Alternative Name(s)

Health Research Board, Ireland, An Bord Taighde Sláinte, HRB

Funding Body Type
Government organisation

Funding Body Subtype
National government

Location
Ireland

Results and Publications

Individual participant data (IPD) sharing plan
The datasets generated during and/or analysed during the current study will be stored in a publicly available repository at https://osf.io/y9hvj/?view_only=c9688c06b34c465994073be1baa9e9a6

• The type of data stored: Survey responses, including comprehension scores, trustworthiness ratings, demographic data
• Dates of availability: Available upon publication of study results
• Whether consent from participants was required and obtained: Informed consent was obtained from all participants, including explicit information about public data sharing
• Comments on data anonymisation: All data are fully de-identified and anonymised: no identifiers collected, demographics provided in ranges only, Prolific payment data handled separately

IPD sharing plan summary
Stored in publicly available repository

Study outputs

Output type	Details	Date created	Date added	Peer reviewed?	Patient-facing?
Protocol article		20/04/2026	16/11/2026	Yes	No
Participant information sheet		11/11/2025	11/11/2025	No	Yes
Study website			16/11/2026	No	No