



Supplementary Figure S1. PubMed landscape and literature-distribution analytics in LipoScope.

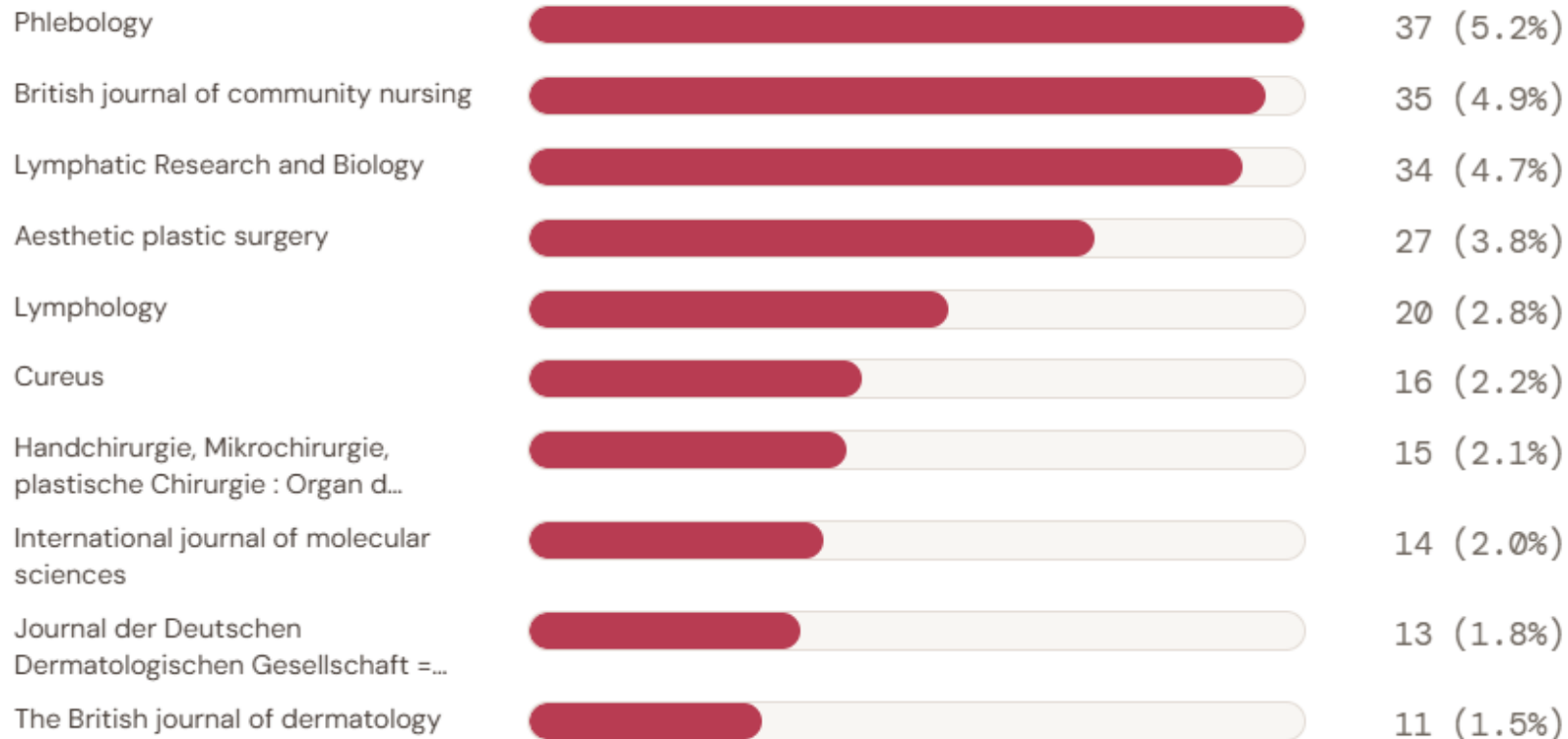
(A) Live corpus summary for the lipoedema/lipedema PubMed search, showing the total number of retrieved records, records published since 2021, unique journals, year of the earliest record and number of papers in the user's Browse Papers workspace. Panels reflect the live evidence snapshot displayed on 28 July 2026.



TOP JOURNALS

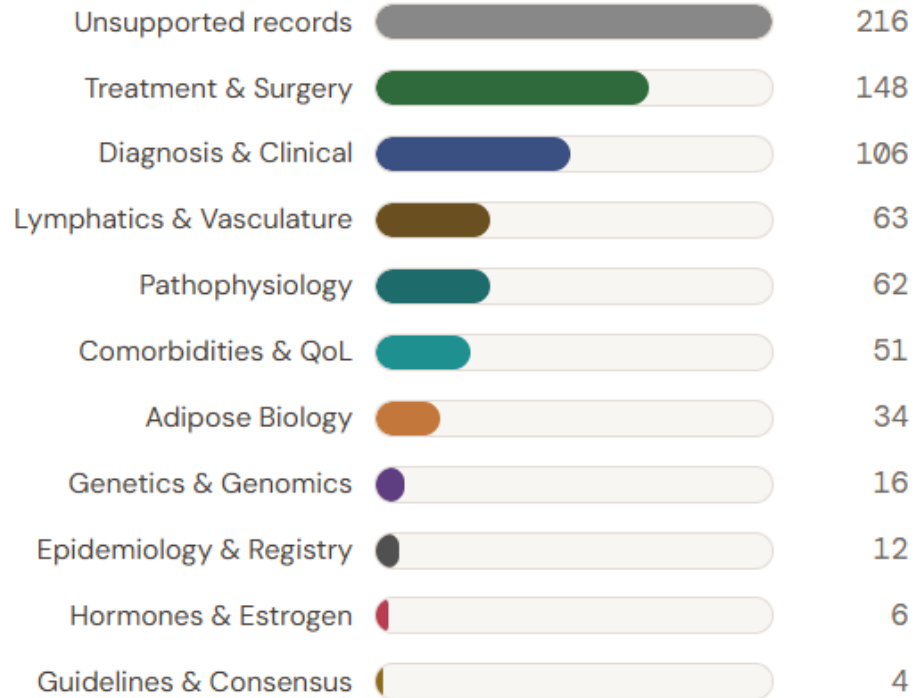
RECORD COUNT AND % OF RECORDS WITH AN IDENTIFIED JOURNAL

Each journal is shown as **record count (percentage)**. Percentage = records assigned to that harmonised journal ÷ all PubMed records with an identified journal × 100. The denominator is **716** identified-journal record(s); 2 record(s) without an identified journal are excluded from the percentage calculation. Counts are refreshed from the live PubMed-derived record set.



(B) Journals with the highest numbers of indexed records. Values are presented as record count and percentage of PubMed records with an identified, harmonized journal; records without journal information are excluded from the percentage denominator. Panels reflect the live evidence snapshot displayed on 28 July 2026.

Primary record-level categorisation: every retrieved PMID is counted once as one primary topic or as Unsupported records. This distribution does not use overlapping query totals.



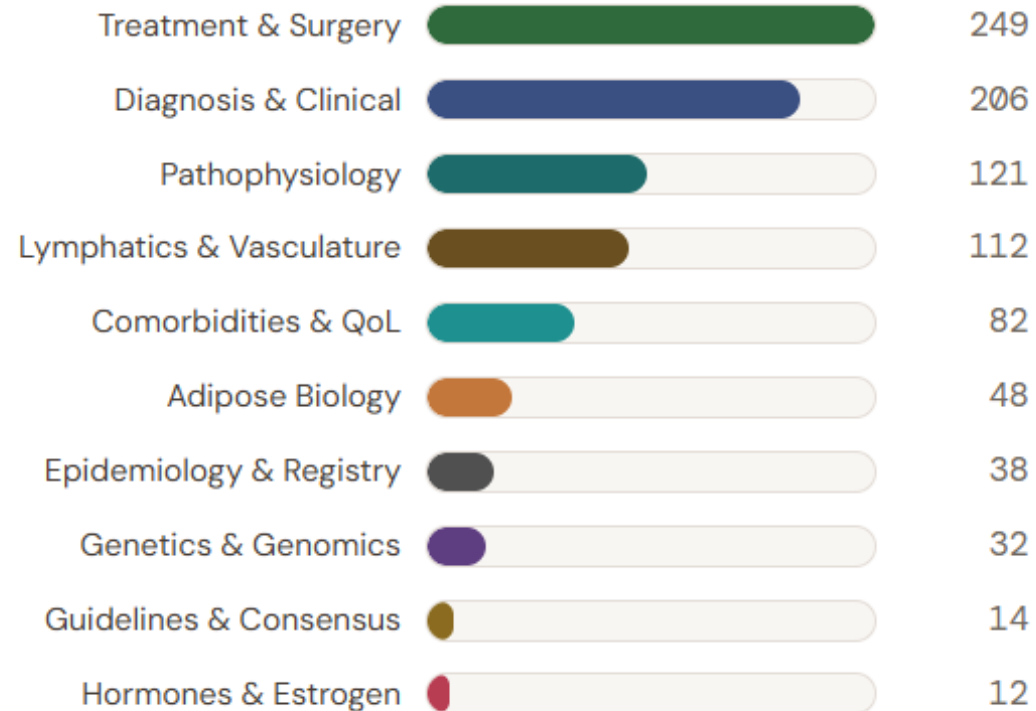
Why records remain unsupported:

Multiple equivalent topic matches: 141 · No category-specific evidence: 30 · PubMed-not-MEDLINE: 26 · Awaiting MEDLINE indexing: 18 · Metadata retrieval incomplete: 1

(C) Mutually exclusive primary topic categorization, in which each classified record is assigned to one principal topic or to *Unsupported records*. The accompanying audit summarizes why some records could not be assigned uniquely, including equivalent topic matches, insufficient category-specific evidence, incomplete MEDLINE indexing and incomplete metadata retrieval. Panels reflect the live evidence snapshot displayed on 28 July 2026.

Exact PubMed ESearch category counts · 28/07/2026, 12:18 pm

PubMed category-query matches: each count comes from the exact query opened by the topic link; categories may overlap.



(D) Independent PubMed ESearch category-query counts generated from the exact search associated with each topic. These categories may overlap because an individual record can satisfy more than one topic-specific query. Panels reflect the live evidence snapshot displayed on 28 July 2026.

LipoScope CCRF Version 1 source-fidelity and implementation workflow

Verified against LipoScope/LipoScope_v1.0.0 and the release-pinned official CCRF XLSX dictionaries
Solid blue = default release build • green = official-source verification/fidelity • purple = browser behaviour

A. Official source and fidelity validation

Official Lipedema Foundation CCRF Version 1 REDCap dictionaries

Four classification-specific XLSX sources are pinned by filename, byte size and SHA-256.

Core: 300 technical rows
Supplemental Highly Recommended: 445
Supplemental: 650
Exploratory: 739

The classifications are cumulative and overlapping.

Corrected source-workbook validation

02_validate_source_workbooks.py verifies the pinned XLSX hashes and counts actual REDCap dictionary records using populated "Variable / Field Name" rows, rather than Excel worksheet dimensions.

Independent official XLSX ↔ frozen TSV fidelity test

09_validate_crf_xlsx_snapshot_fidelity.py reparses the four exact pinned XLSX dictionaries and independently regenerates the canonical row-level TSV. It compares every retained source/provenance column, row order and source-row position with the frozen snapshot used by LipoScope.

PASS: 2,134/2,134 rows matched; 0 row differences; 0 field differences.

Byte-identical validated frozen source snapshot

The independently regenerated TSV and the committed release snapshot have the same byte content and the same SHA-256:

74934b25105072e33e949e2f145eee2e09435028b5cdd528424a55c5de79046b
data/ccrf_source_snapshot.tsv

What the fidelity result establishes

The frozen TSV is a reproducible file-format representation of the pinned official CCRF dictionaries. The tested questions, field definitions, coded choices, validation metadata, branching logic and other retained source metadata are preserved without detected differences.

B. Default deterministic release build

Default build inputs

The release build uses the validated frozen source snapshot together with official counts and source configuration:

data/ccrf_source_snapshot.tsv
data/ccrf_official_counts.json
config/ccrf_sources.yaml

1. Build the CCRF application schema

03_build_crf_schema.py reads the validated frozen source representation, retains source field metadata, derives the display section within each form and writes data/ccrf_schema.json.

2. Reconcile frozen source rows with generated schema

04_compare_schema_to_sources.py compares the generated schema row-by-row against the frozen snapshot for the explicitly tested source attributes.

Release result: 2,134/2,134 technical rows matched; 0 differed.

3. Validate official counts and calculated-field policy

06_validate_count_alignment.py keeps official CDE totals separate from technical rows and completion denominators. 07_validate_calculated_fields.py confirms that no formula is inferred when an official executable expression is absent.

4. Embed the validated schema into LipoScope

08_embed_crf_schema.py serialises data/ccrf_schema.json into the application and writes matching template, active and versioned release HTML copies.

Release validation

The committed release records 35 automated tests passed, browser audit passed, Python syntax validation passed, static HTML validation passed and deterministic schema/HTML-copy validation passed.

C. Browser implementation

Interactive CCRF forms

The embedded schema drives ordinary browser controls. Users choose one of four CCRF classifications and complete source-defined fields. Official coded choices and source validation are retained where supplied.

Conditional questions

Official branching expressions determine whether a question is currently applicable. Inactive responses can be retained and identified rather than silently deleted.

Completion and data-quality review

LipoScope assesses currently applicable required fields that can be evaluated locally and identifies missing, invalid, inactive-branch and version-related issues. This workspace completion measure is distinct from the official CDE total.

Direct identifiers

Source-marked identifier fields remain represented in the data dictionary but are not populated in persistent browser-local records or research-data exports.

Research package

Users can prepare coded, human-readable, code-and-label and long-format records together with the data dictionary, data-quality report and source manifest.

Key distinction: official XLSX ↔ frozen TSV fidelity is independently validated; the default deterministic build then starts from that validated frozen snapshot.

Supplementary Figure S2: LipoScope CCRF Version 1 source-fidelity and implementation workflow. Overview of the LipoScope Common Case Report Form (CCRF) implementation, showing validation of the official Lipedema Foundation CCRF Version 1 REDCap dictionaries, deterministic generation and reconciliation of the embedded application schema and its use in browser-based data entry, branching, completeness assessment and research-data export. Source-fidelity testing confirmed that all 2,134 cumulative technical rows were reproduced at the source/schema level with no detected row or field-attribute differences; this comparison does not imply identical reproduction of native REDCap presentation or visual features, including matrix grouping or matrix-level instructional display.

Supplementary Note S1

LipoScope v1.0.0

User Guide

Literature discovery, PubMed monitoring, analytics and CCRF-based standardized research data collection

Application scope: LipoScope is a browser-local research workbench. It supports literature discovery, project-specific paper/article organisation, PubMed-wide analytics and de-identified research-data preparation using the Lipedema Foundation Common Case Report Form Version 1. It is not a diagnostic system and does not replace an institutionally governed data-capture platform.

Application version	v1.0.0
Application update	3 August 2026
Hosted application	Open LipoScope
Source repository	View source repository
Guide audience	Clinicians, nurses, researchers, patient organisations

Contents

- [1. Purpose and intended use](#)
- [2. Quick start](#)
- [3. Interface and browser-local workspace](#)
- [4. Start Here](#)
- [5. Browse Papers](#)
- [6. Live Search and New Paper Alerts](#)
- [7. Analytics](#)
- [8. CCRF Workspace](#)
- [9. About and workspace management](#)
- [10. Recommended workflows](#)
- [11. Privacy, governance and safe use](#)
- [12. Troubleshooting](#)
- [13. Glossary](#)
- [Appendix A. Quick-guide for users](#)

Using this guide: Interface labels are shown in bold and retain the wording used in LipoScope. The guide separates functions that use live internet services from functions that remain local to the current browser.

1. Purpose and intended use

LipoScope is designed to reduce the practical distance between finding lipoedema literature and preparing structured research data. It provides a disease-focused literature workspace, current PubMed searching, literature analytics and an interactive implementation of the Lipedema Foundation Common Case Report Form Version 1.

The application is suitable for users who need to locate relevant publications, maintain a project-specific reading set, compare their paper collection with the wider PubMed landscape, inspect official Common Case Report Form definitions, complete de-identified research records and export standardised data for later analysis or governed institutional use.

LipoScope can help with	LipoScope does not do
Finding and organising lipoedema papers	Diagnosing lipoedema or determining clinical eligibility
Monitoring newly indexed PubMed records	Replacing critical appraisal of original publications
Reviewing structured article summaries and available figures	Guaranteeing full-text or figure access for every paper
Comparing live PubMed trends with a local paper collection	Treating the 200-paper starting library as a systematic-review corpus
Preparing de-identified CCRF-aligned research records	Storing direct privacy identifiers or signatures
Checking missing or invalid required CCRF responses	Replacing REDCap or another governed institutional data system

Research-use: Use LipoScope for literature discovery, research planning, de-identified data organisation and completeness review.

1.1 Who may use LipoScope

User group	Typical use
Researchers	Curate papers, identify topic gaps, prepare CCRF-aligned data and export analysis-ready files.
Clinicians and nurses	Locate evidence, review article summaries and figures, inspect and complete CCRF questions and use readable exports for research review.
Patient organisations	Monitor publications, organise disease-specific resources and support research-planning.
Students and trainees	Learn the literature landscape and understand standardised research-data structures.

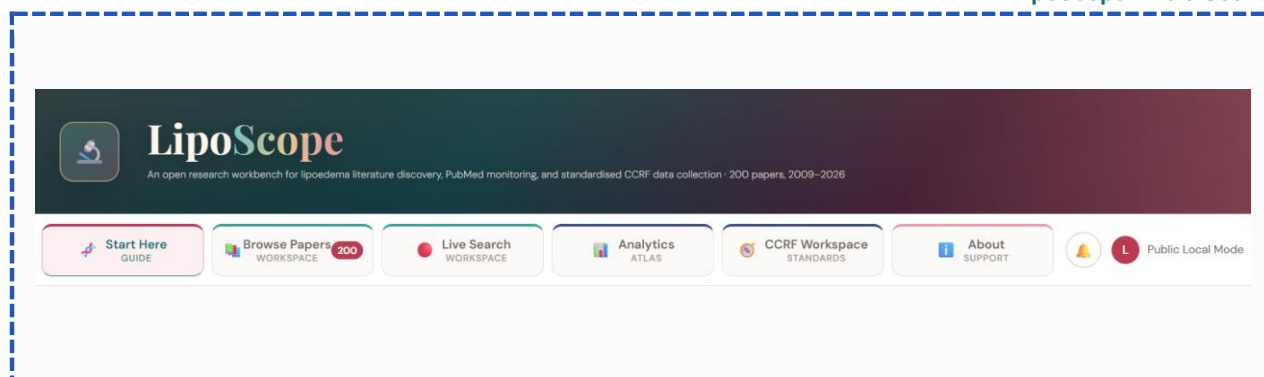


Figure 1. LipoScope application overview

2. Quick start

- 1 **Open LipoScope.** Use the hosted application or open the downloaded LipoScope.html file in a current desktop browser.
- 2 **Review the Start Here tab.** Use the five-step “Guided App Walkthrough” to understand the main literature and CCRF workflows.
- 3 **Open Browse Papers.** Search or filter the 200-paper starting library and open a paper card.
- 4 **Use Live Search when needed.** Search PubMed for a paper that is not in the current Browse workspace and add it locally.
- 5 **Open Analytics.** Compare the live PubMed landscape with the papers currently visible in Browse Papers.
- 6 **Open CCRF Workspace.** Choose a CCRF classification, create a de-identified record, complete applicable fields and review the Gap Radar.
- 7 **Export and back up.** Share literature or export CCRF outputs when required. Browser-local data may be lost if browser storage is cleared.

2.1 Technical requirements

- **Browser:** Use a current desktop browser. The interface may be less practical on a small mobile screen.
- **Internet connection:** Required for PubMed searches, new-paper checks, live analytics, Google Scholar links, metadata refresh and open-access figure retrieval.
- **Local functions:** Visible-paper filters, hidden-paper preferences, user-added paper records, persistent local PDF previews and saved CCRF records are stored in the current browser.

- **Downloads:** Allow browser downloads and pop-up windows where needed for PubMed, DOI, Google Scholar and exported files.
- **Backups:** Export important CCRF records and research packages regularly; browser-local storage is not a managed backup system.

Hosted versus downloaded use: The hosted version is generally the simplest option for live internet functions. When a downloaded HTML file is opened directly, browser security settings may affect external requests; try the hosted version or another current browser if live services do not load.

3. Interface and browser-local workspace

The top navigation bar contains six principal tabs. A notification bell provides new-paper alerts, and the profile area opens the About tab for optional local workspace labels and sharing controls.

Navigation item	Purpose
Start Here	Orientation, guided walkthrough and links to research, clinical and community resources.
Browse Papers	Search, filter, review and curate the current paper collection.
Live Search	Search PubMed or Google Scholar, run topic-based discovery and add PubMed records.
Analytics	Compare live PubMed-wide literature landscape with the currently visible Browse collection.
CCRF Workspace	Create and review de-identified CCRF-aligned records and export standardised data.
About	Review scope, citation, privacy, topic definitions, local profiles, sharing and database-management controls.
New Paper Alerts bell	Check recent PubMed records, open articles and add new articles to Browse Papers.
Profile area	Open local profile and sharing controls. A profile label is not an account or login.

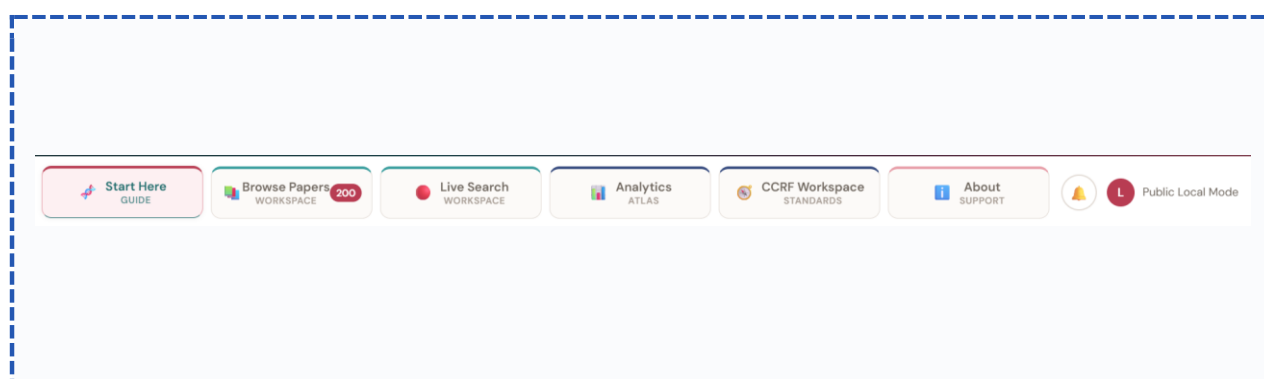


Figure 2. Top navigation and workspace controls

3.1 What “browser-local” means

Browser-local means that LipoScope stores workspace state in the browser being used. Data are not automatically synchronised to another browser, device or institutional database. A different browser, private-browsing session or cleared site storage may show a different workspace.

Stored locally	Not automatically included elsewhere
Hidden-paper choices	Not transferred to another device unless a Browse share link is used.
User-added PubMed paper cards	Not part of the distributed application or PubMed-wide analytics denominator.
Optional persistent local PDF previews	Not uploaded to GitHub and not included in shared Browse links.
CCRF records	Not included in Browse share links and not transmitted to a LipoScope server.

Local profile labels: A local profile label separates collections within the same browser. It is not authentication, does not protect data from other users of the same browser account and should not be treated as a security control.

4. Start Here

The Start Here tab introduces the application in plain language. It explains the relationship between the starting literature library, live PubMed functions, Analytics and the CCRF Workspace. It also displays a live PubMed status message when internet-derived literature information is available.

A Simple, Browser-Local Workspace for Lipoedema Research

LipoScope brings together a starting library of 200 lipoedema papers, current PubMed searching, field-level analytics and the official CCRF Version 1 research workspace. Use it to find relevant papers, understand what has been studied, organise a focused reading set and prepare standardised research records. Your hidden papers, added papers, notes, local PDF figure previews and CCRF records remain in this browser unless you choose to share or export them. LipoScope supports research planning and data organisation; it does not diagnose a patient or replace an institutionally governed data-capture system.

Find relevant papers

Search the starting library by topic, study type, year, title, author, journal or findings, and add recent PubMed records through Live Search.

Keep a focused literature set

Hide papers that are not relevant, restore them later, add notes and keep optional PDF figure previews on the same device.

See the wider research field

Compare live PubMed publication trends with the papers currently visible in your Browse workspace.

Prepare standardised research records

Use the official CCRF Version 1 classifications, review required-field gaps and export coded or readable datasets.

Work locally

Most workspace changes remain in your browser unless you deliberately share a Browse link or download an export.

PubMed literature view checking PubMed automatically · 718 PubMed records cached · last attempt 05/08/2026, 10:12 pm

Figure 3. Start Here overview

4.1 Guided App Walkthrough

The walkthrough contains five slides. Select Previous or Next to move manually, or select Play for an automatic tour. The progress bar and step number show the current position.

Walkthrough step	What it teaches
1. Find and review papers	Search and filter Browse Papers, open article cards and locate source links and figures.

Walkthrough step	What it teaches
2. Add recent PubMed records	Search Live Search, review a citation and add a record to the local Browse collection.
3. Understand the research landscape	Interpret the separate live PubMed and visible-workspace Analytics sections.
4. Prepare a CCRF research record	Choose a classification, complete applicable fields, review gaps and export data.
5. Manage your local workspace	Hide or restore default papers, understand local storage and use deliberate sharing.

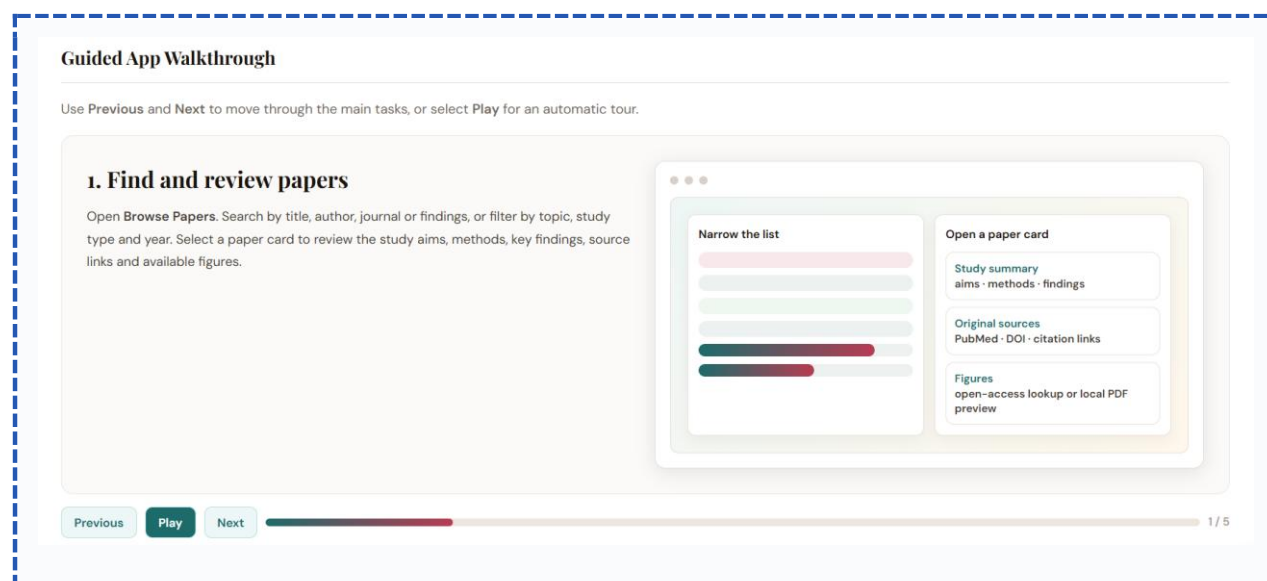


Figure 4. Guided App Walkthrough

4.2 External resources

The lower Start Here section links to research, clinical and community resources. These links open external websites and are not stored or controlled by LipoScope.

Resource	Use
Lipedema Foundation	Research, education, registry and disease resources.
Legato Lipedema Research Library	A specialised literature archive that can complement PubMed-based discovery.
Lipedema Research Roadmap	Research priorities and identified field gaps.
Lipedema Foundation Registry	Registry information and patient-research participation.
US Standard of Care for Lipedema	Open-access consensus clinical guidance.
ClinicalTrials.gov	Current interventional and observational studies.
Lipoedema Australia	Australian education and support resources.
Legs Matter	Public lower-limb health information.
LE&RN Resource Center	Lymphatic disease education and support.

Resource	Use
NCBI MedGen Lipedema	Indexed biomedical terminology and linked resources.

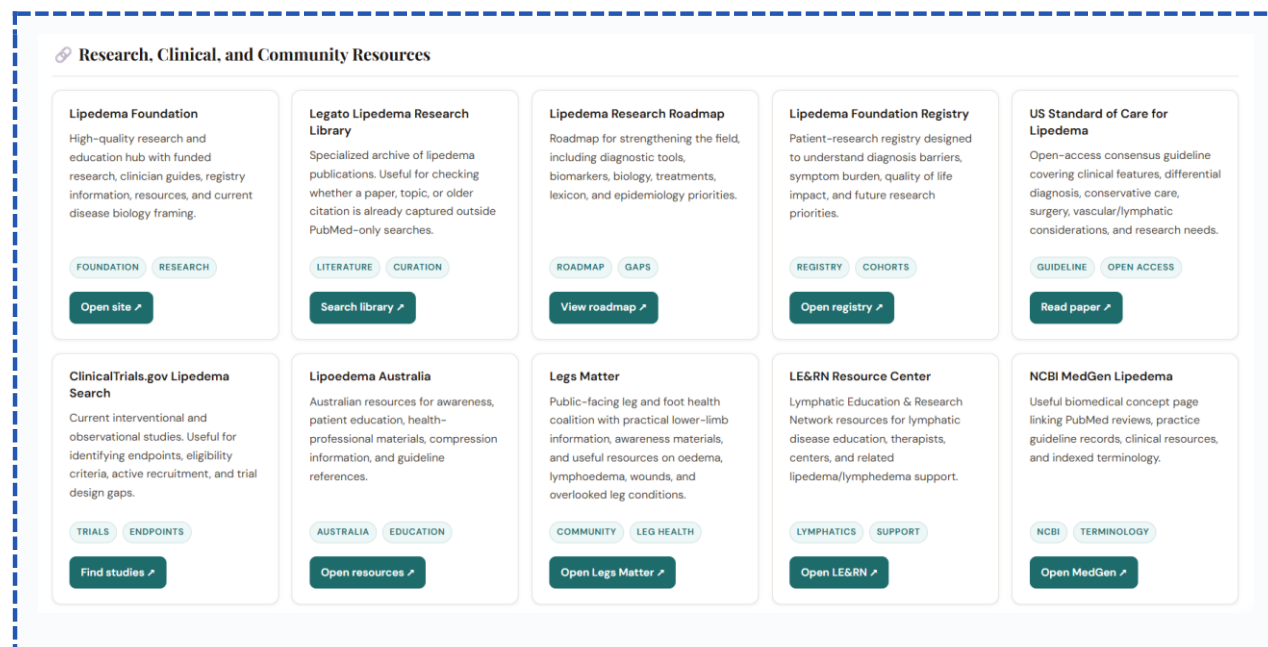


Figure 5. Research, clinical and community resources

5. Browse Papers

Browse Papers is the project-specific literature workspace. The distributed version contains a preconfigured 200-paper starting library so that users can begin exploring the interface immediately. This library is an orientation and demonstration collection, not a systematic-review corpus and not the denominator used for PubMed-wide Analytics.

Starting library: You may keep the 200 papers, hide selected papers or hide the entire starting library through About. User-added records are stored separately in the current browser and can be removed without affecting the distributed papers.

5.1 Search and filters

Control	Function
Search box	Searches visible records across title, authors, journal, summary and findings.
Topic filter	Limits records to one displayed primary topic, including Unsupported records.
Study type filter	Limits records to a study-design or publication-format category, including Unclassified / Other.
Year filter	Limits records to a selected year or 2017 and earlier.
Clear filters	Returns the visible collection to the unfiltered current workspace.

Control	Function
Results count	Shows the number of papers matching the filters and the total number visible in the workspace.

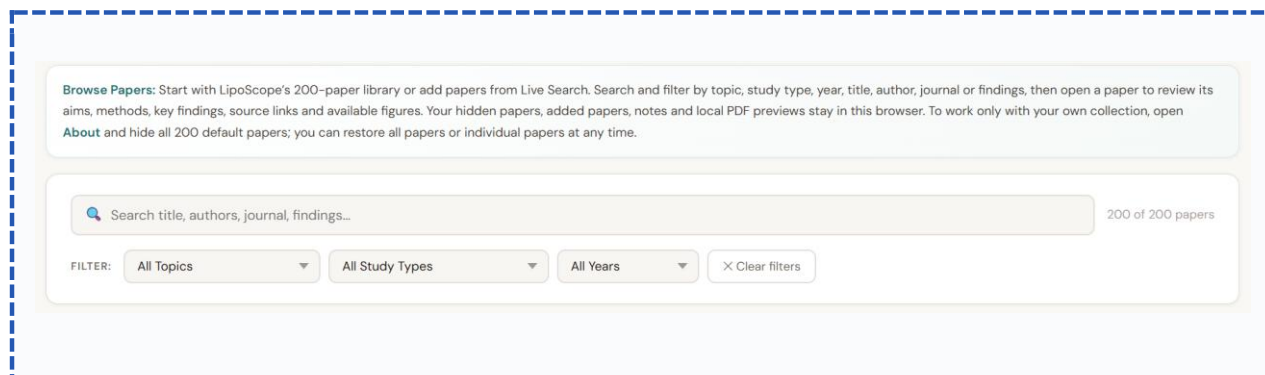


Figure 6. Browse Papers search and filter controls

5.2 Understanding an article card

Each card summarises one paper. The top badges show the year, study type and primary topic. User-added records carry an Added badge, and papers with a stored local PDF preview may show a PDF badge. The card then displays the title, authors, journal and a short structured summary.

Card action	What happens
Select the card	Opens the detailed paper window.
Full Text	Opens the DOI or available source link in a new tab.
Figures	Opens the paper window and starts figure lookup.
Hide	Removes a distributed paper from the visible workspace but allows later restoration.
Remove	Deletes a user-added paper from the current browser workspace.

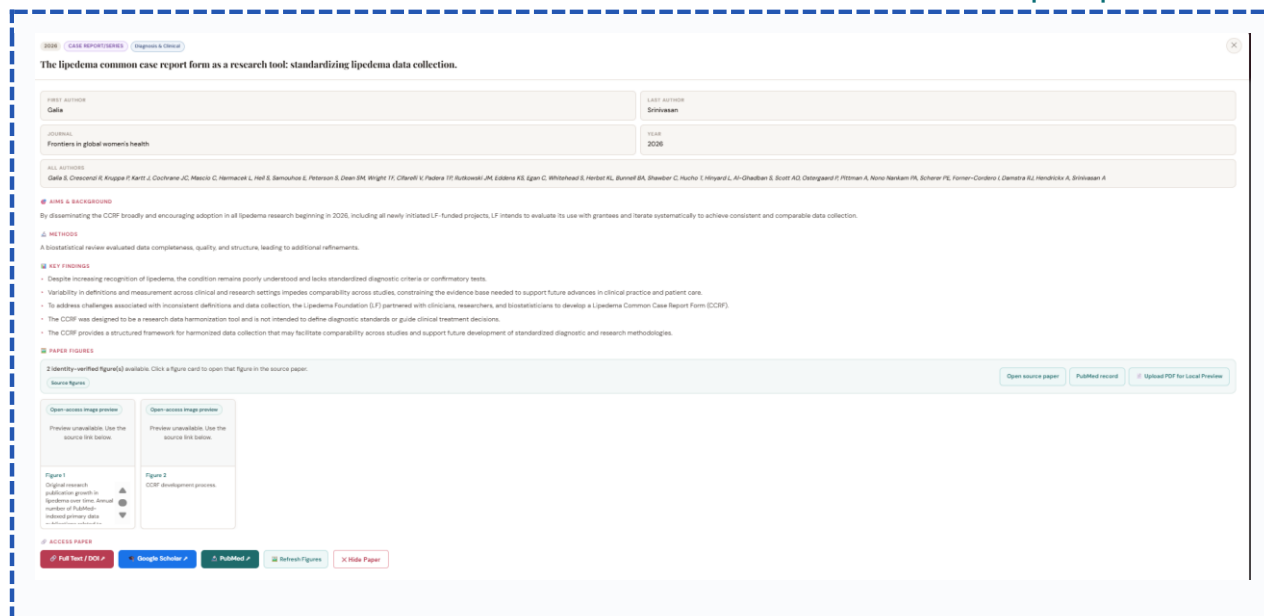


Figure 7. Anatomy of a Browse Papers article card

5.3 Detailed paper window

Opening a paper displays the available metadata and structured sections. Depending on the source record, the window may include Aims and Background, Methods, Key Findings, author and journal details, topic information, figure access and external source links. Structured summaries support navigation and should be checked against the original publication before being used in research interpretation.

Section	Content
Metadata badges	Publication year, study type, primary topic and any supported secondary topic or unresolved-classification reason.
Author and journal information	Available authorship, journal and citation details.
Aims and Background	Purpose and context derived from available source information.
Methods	Available study-design or methods information.
Key Findings	Concise findings extracted or structured from available abstract information.
Paper Figures	Open-access figure lookup and optional local PDF preview.
Access Paper	DOI/full text, Google Scholar and PubMed links, plus figure refresh and hide/remove controls.

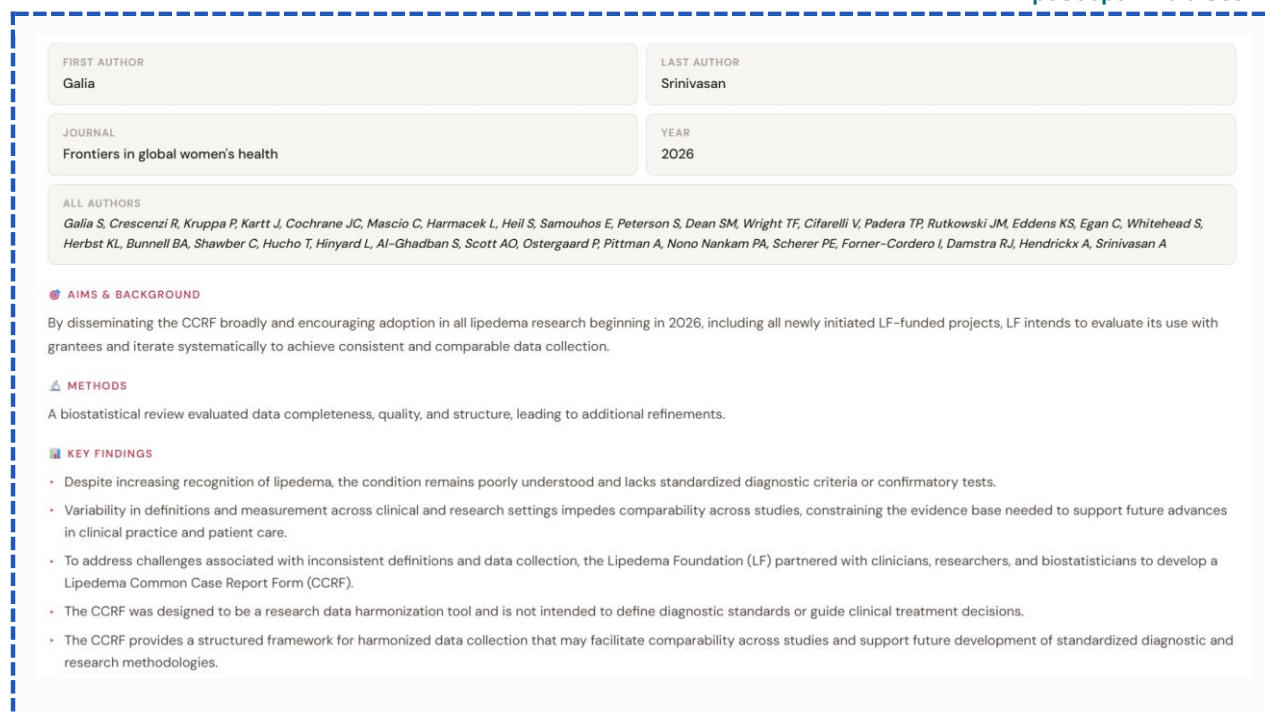


Figure 8. Detailed paper window

5.4 Figure access

LipoScope first attempts to identify figures from an identity-matched open-access PubMed Central record. Not every paper has an accessible PubMed Central full text, and the absence of figures in LipoScope does not mean that the publication contains no figures.

When open-access figure retrieval is unavailable or incomplete, a user may select a local PDF for device-local figure preview. The PDF is used only in the current browser. If persistent local PDF storage is enabled in About, the PDF can be kept on the device for the current local workspace; otherwise, the preview is session-only and may disappear after refresh or page closure.

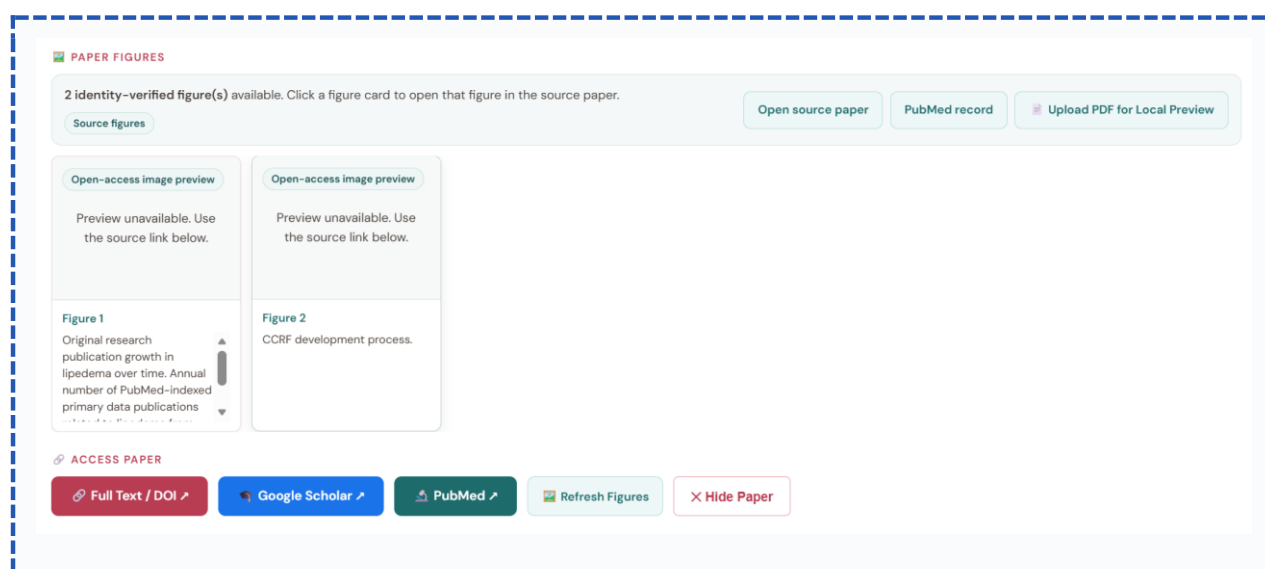


Figure 9. Paper figure panel and local PDF preview

5.5 Hiding, removing and restoring papers

- **Hide a distributed paper:** Select Hide on the article card or in the paper window. The paper remains part of the application but is removed from the visible workspace.
- **Remove a user-added paper:** Select Remove. This deletes the locally added record from the current workspace.
- **Restore one hidden paper:** Open About and use the individual Hidden Papers list.
- **Restore all hidden papers:** Open About and select Restore All Hidden Papers.
- **Hide the full starting library:** Open About and select Hide All Default Literature.
- **Restore the original starting library:** Open About and select Restore 200 Default Corpus.

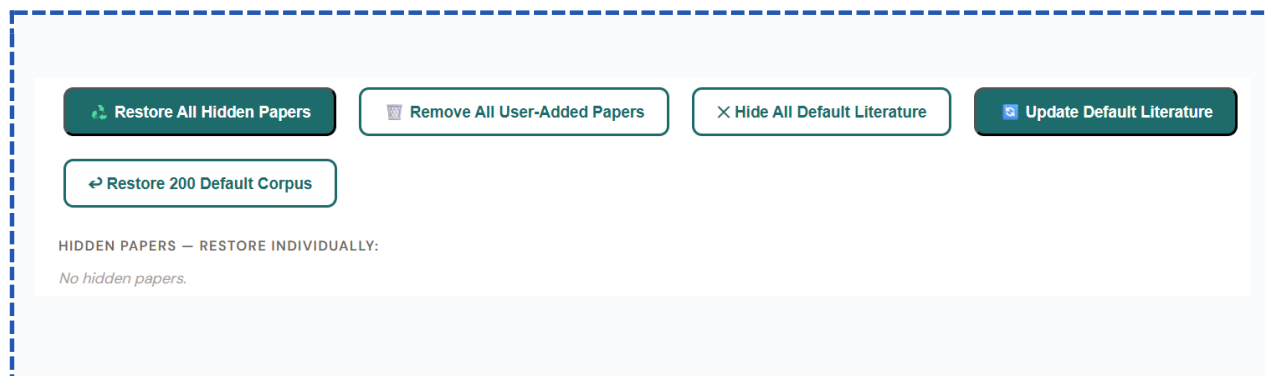


Figure 10. Hidden-paper and restoration controls

6. Live Search and New Paper Alerts

Live Search adds current literature discovery to the browser-local workspace. Search results are fetched from external services and are not stored unless a PubMed record is added to Browse Papers.

6.1 PubMed search

Select PubMed, enter an author, gene, treatment, method or research phrase and select Search. LipoScope automatically combines the query with the lipoedema/lipedema disease scope and returns up to 20 recent PubMed results. Each result displays the title, year, authors, journal, PMID and available source links.

1

Enter a search term. Examples include AKR1C1, lymphatic imaging, pain, liposuction or a known paper title.

2

Select PubMed. The query is restricted to PubMed records that also match the lipoedema/lipedema disease scope.

3

Review the result. Open PubMed, the DOI/full text or Google Scholar if additional source checking is needed.

4

Add a useful paper. Select + Add to Browse. The citation and available abstract-derived summary are stored in the current browser.

5

Return to Browse Papers. The added paper appears as a user-added article card and affects only the visible-workspace Analytics section.

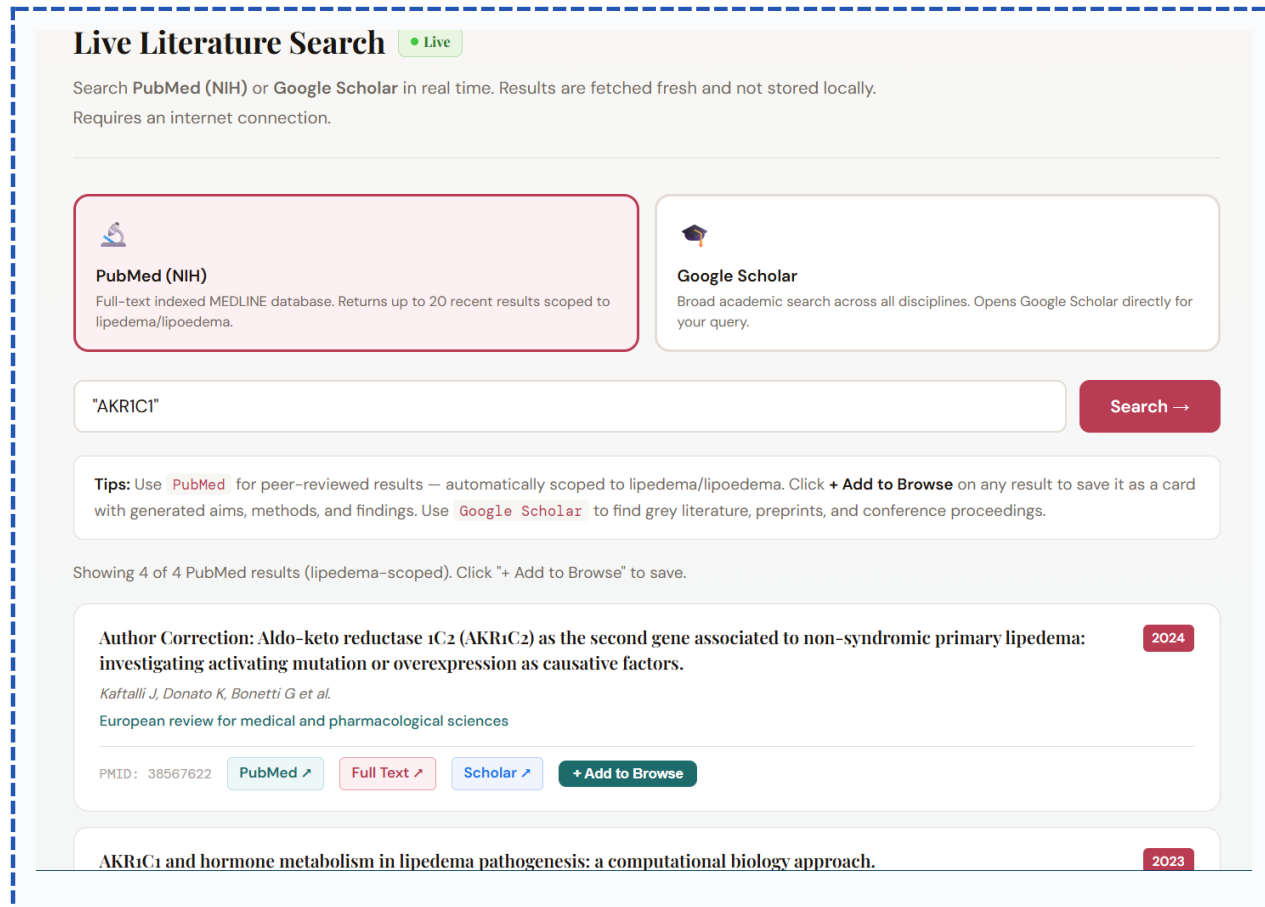


Figure 11. Live PubMed search

6.2 Google Scholar search

Select Google Scholar to open a broader external scholarly search. LipoScope adds lipoedema/lipedema terms to the search and opens Google Scholar in a new browser tab. Google Scholar results are not imported into LipoScope and do not contribute to PubMed-wide Analytics.

When to use Google Scholar: Google Scholar can help locate grey literature, theses, conference material, preprints or alternate full-text locations. Verify publication status and provenance before using any result in research.

6.3 Auto-Discovery Engine

The Auto-Discovery Engine searches the configured lipoedema research topics through PubMed. Topic terms are accepted only when the returned record also matches the disease scope. Records are deduplicated, classified and imported into the user-added collection.

Control	Meaning
Max per Topic	Sets the maximum number of recent records requested for each selected topic: 10, 20, 30 or 50.
Year From	Restricts discovery to any year or records from 2020, 2022, 2023, 2024 or 2025 onward.
Topics to Search	Allows topic-by-topic selection, with Select/Deselect All.
Discover & Import Papers	Runs the selected PubMed searches and imports new disease-relevant records.
Stop	Stops the discovery run after the current operation completes.
Remove Duplicates	Compares user-added papers with the default corpus using PMID and normalised title, removes duplicate local records and keeps the distributed record.

Review after auto-discovery: Auto-Discovery can add multiple records without individual preselection. Review the resulting Browse cards and remove records that are not relevant to the project.

Auto-Discovery Engine Auto

Automatically search PubMed across all 10 research topics, while enforcing a strict **lipedema/lipoedema disease scope**. Topic terms such as lymphatic, adipose, hormone, or fibrosis are only accepted when the record also discusses lipedema/lipoedema. Results are classified from title + abstract text and deduplicated before import.

SEARCH SOURCE: PubMed-only discovery

MAX PER TOPIC: 20 papers

YEAR FROM: Any year

TOPICS TO SEARCH Select/Deselect All

- ☒ Genetics & Genomics
- ☒ Pathophysiology
- ☒ Hormones & Estrogen
- ☒ Diagnosis & Clinical
- ☒ Treatment & Surgery
- ☒ Lymphatics & Vasculature
- ☒ Adipose Biology
- ☒ Epidemiology & Registry
- ☒ Guidelines & Consensus
- ☒ Comorbidities & QoL

Discover & Import Papers Remove Duplicates

Remove Duplicates: scans user-added papers against the built-in/default corpus by PMID and normalized title, removes duplicate local records, keeps the curated default record, then refreshes Browse and visible-corpus Analytics. Live Search and Analytics remain unaffected by local duplicate removal.

Figure 12. Auto-Discovery Engine

6.4 New Paper Alerts

The bell in the top navigation opens the New Paper Alerts panel. Select Check Now to search recent PubMed records that are not already represented in the current database. The last-check time is stored locally, and repeated automatic checks may be throttled to avoid unnecessary requests.

Alert control	Function
PubMed	Opens the record in PubMed.
Add to Browse	Retrieves the citation and abstract information and creates a local article card.
Mark all read	Clears the unread status without deleting alert items.
Clear All	Deletes the current alert list from the browser.
Check Now	Forces a current recent-record check.
Close	Closes the alert panel.

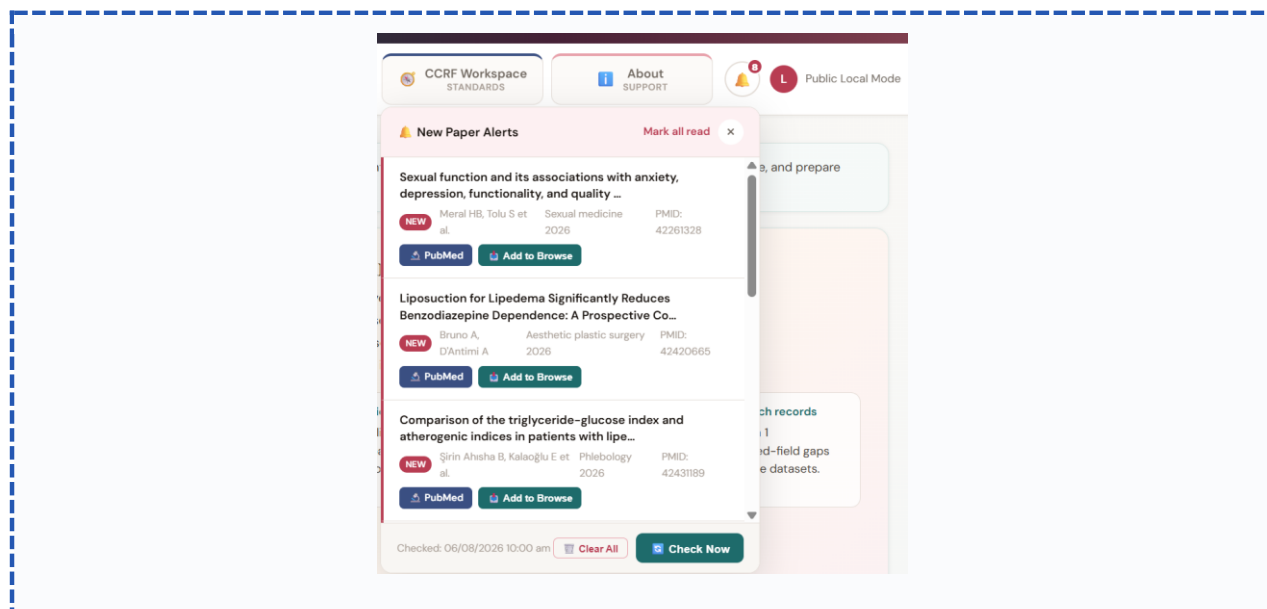


Figure 13. New Paper Alerts

7. Analytics

Analytics deliberately separates two denominators. The upper section represents the wider live PubMed lipoedema/lipedema literature. The lower section represents the papers currently visible in Browse Papers. These views should be compared, not combined.

Live counts can change: PubMed indexing, metadata and newly published records change over time. Record the query date when reporting values from the live Analytics tab.

7.1 Live PubMed landscape

The live section uses one canonical disease query and paginated PubMed retrieval. If complete data cannot be retrieved and no successful cache is available, LipoScope displays an unavailable message rather than substituting estimated counts.

Display	How to interpret it
Summary statistics	Current total records, recent records, identified journals, earliest record and related live measures.
Research Topics: query matches	Exact topic-specific PubMed query counts. A paper may match more than one category.
Research Topics: primary categorisation	Each retrieved record is counted once as one primary topic or Unsupported records.
Publication Timeline	Annual record counts; records without a resolved year are reported separately.
Study Types	Classification based first on PubMed/NLM metadata, followed by verified full-text and design evidence where needed.
Topic Cloud	Clickable category chips that open the exact PubMed topic query.
Top Journals	Record count and percentage among records with an identified harmonised journal.

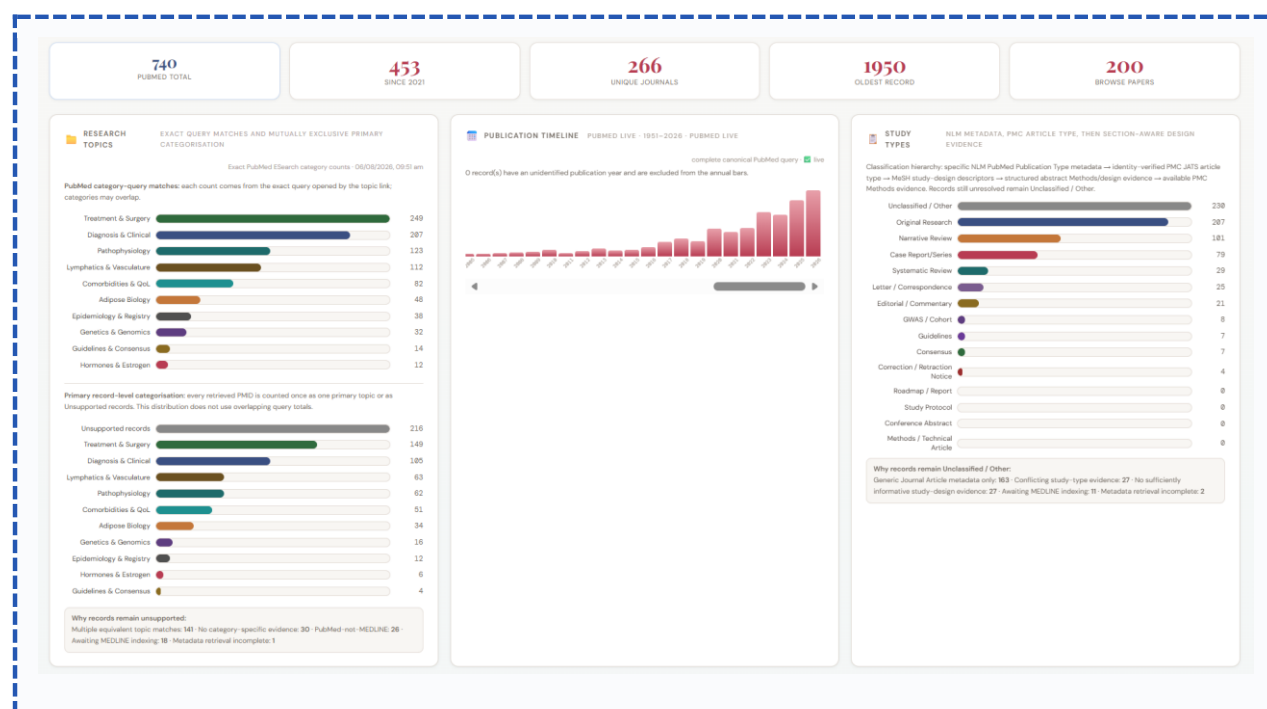


Figure 14. Live PubMed Analytics summary

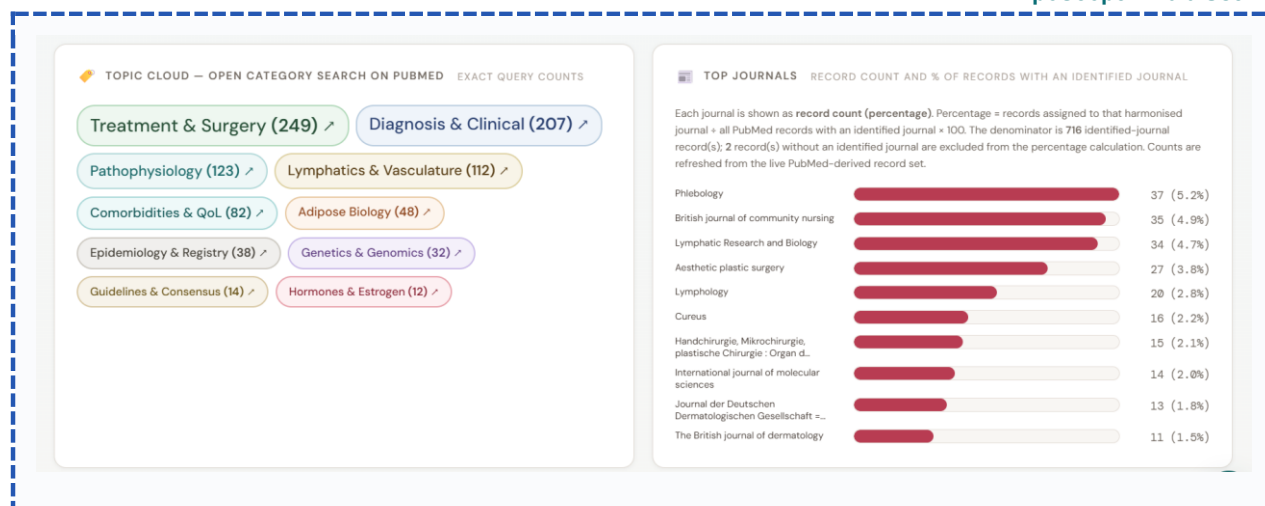


Figure 15. Topic Cloud and Top Journals

7.2 Visible LipoScope corpus

The Visible LipoScope Corpus charts update immediately when the user filters, adds, hides, restores or removes papers. The topic distribution and year distribution therefore describe the current working collection, not the whole PubMed field.

- **Use the comparison to identify collection bias:** A genetics-focused project may intentionally over-represent genetics papers compared with the whole field.
- **Check the visible count:** The displayed number reflects the current Browse workspace after hiding and adding records.

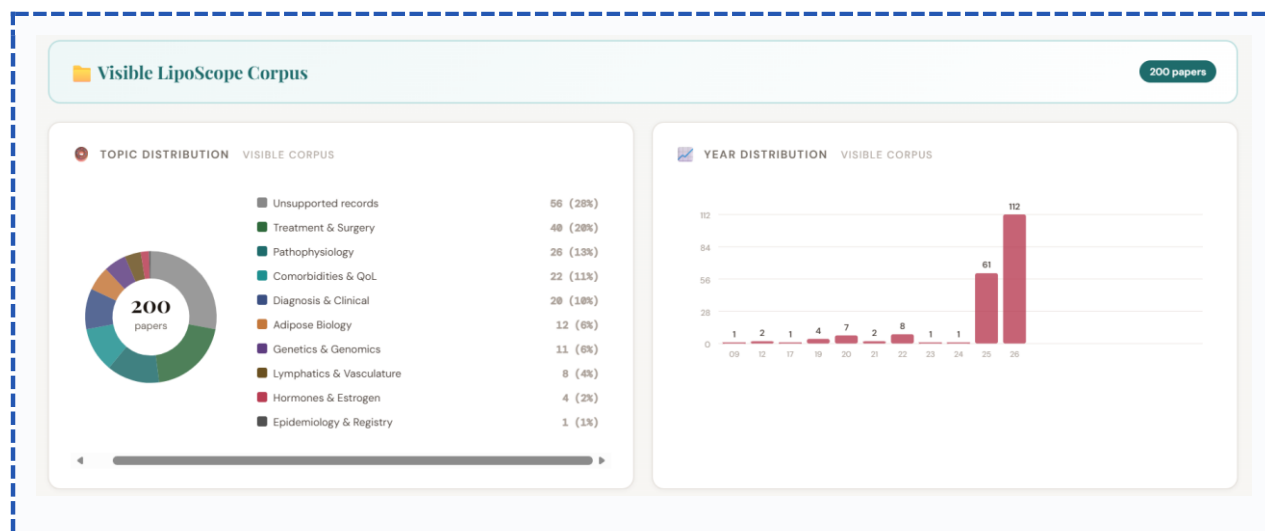


Figure 16. Visible-workspace Analytics

7.3 Topic and study-type cautions

Topic-query counts may overlap, whereas primary record-level categorisation is mutually exclusive. A topic count from the Topic Cloud may therefore be higher than the number of records assigned that topic as their primary category. Records lacking sufficient evidence may remain Unsupported records, and records with unresolved study design may remain Unclassified / Other. These labels preserve uncertainty rather than forcing a classification.

8. CCRF Workspace

The CCRF Workspace converts the Lipedema Foundation Common Case Report Form Version 1 data dictionaries into interactive browser forms. The three connected modules share the same selected classification, active record, source definitions and code-to-label mappings.

Module	Purpose
Standardized Cohort Builder	Create a record and complete applicable participant and clinician fields.
Cohort Design Gap Radar	Review missing, invalid, privacy-excluded, inactive-branch and version-mismatch states.
CCRF Data Dictionary and Dataset Export	Inspect official field definitions and export coded, readable and quality-control files.

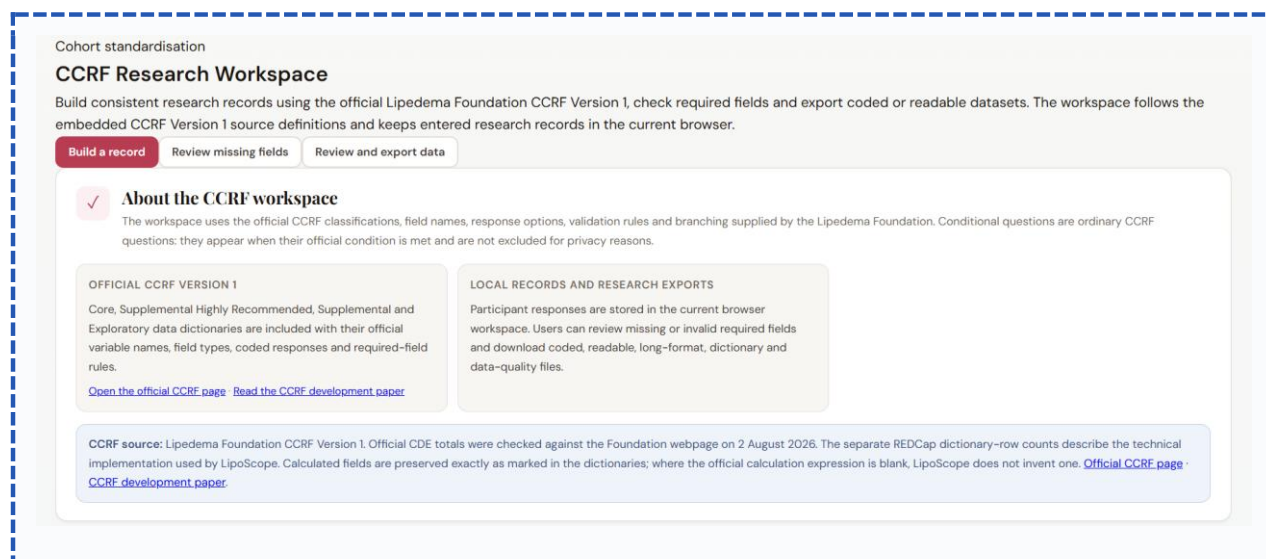


Figure 17. CCRF Workspace overview

8.1 Choosing a CCRF classification

The four classifications are cumulative. Select the classification required by the study protocol or data-collection plan. The official CDE total is distinct from the number of technical REDCap dictionary rows and from the number of currently applicable required response fields in a particular record.

Classification	Official CDEs	Clinician-completed	Participant-completed	REDCap data-dictionary rows
Core	288	86	202	300
Supplemental Highly Recommended	424	89	335	445
Supplemental	618	89	529	650
Exploratory	682	120	562	739

Why the counts differ: REDCap data-dictionary rows can outnumber official CDEs because they also include items needed to build and run the electronic form, such as instructions, identifiers and calculated fields. A record is considered complete based only on the required fields for that specific record, not all CCRF CDEs published in the system.

8.2 Starting and navigating a record

- 1** **Select the official classification.** Choose Core, Supplemental Highly Recommended, Supplemental or Exploratory.
- 2** **Create or continue a record.** Select New record for a fresh generated local record ID, or load a saved record from the ledger.
- 3** **Choose the official form.** Navigate between clinician and participant forms available in the selected classification.
- 4** **Choose the official section.** Select the relevant section within the form.
- 5** **Find a question.** Search by the official question wording, variable name, note or response option.
- 6** **Use display filters.** Show Required only or Missing or invalid only when reviewing a long form.
- 7** **Enter responses.** Use the browser controls generated from the embedded CCRF schema and coded choices, noting that some native REDCap presentation features, including matrix grouping, are represented differently in LipoScope v1.0.0.
- 8** **Save locally.** Select Save record locally to place a sanitised de-identified copy in the current browser profile.

CRF Standardized Cohort Builder
Select an official CCRF classification and complete its source-defined participant and clinical fields. LipoScope shows required-field completion, missing or invalid responses, and saved-record summaries.

Research-use note: use a study identifier for browser-local records and enter identifiable information in the study's governed data-capture system. LipoScope does not diagnose a patient.

OFFICIAL CCRF CLASSIFICATION
Select an official CCRF classification New record Save record locally

Exploratory (Complete CCRF): 682 official CDEs (120 clinician-completed and 562 participant-completed). LipoScope uses 739 REDCap dictionary rows to preserve the questions together with instructions, identifiers, calculated-field definitions and conditional logic. Conditional questions are included normally when their official branching conditions are met. Completion below is calculated only from applicable required response fields that can be assessed in this browser; it is not a percentage of all 682 CDEs.

Select a CCRF classification
Choose Core, Supplemental Highly Recommended, Supplemental, or Exploratory. LipoScope will render the exact fields, coded choices, required flags, and branching rules from that official data dictionary.

Saved CCRF records
No de-identified records have been saved in this browser profile.

Figure 18. Starting a CCRF record

CRF Standardized Cohort Builder
Select an official CCRF classification and complete its source-defined participant and clinical fields. LipoScope shows required-field completion, missing or invalid responses, and saved-record summaries.

Research-use note: use a study identifier for browser-local records and enter identifiable information in the study's governed data-capture system. LipoScope does not diagnose a patient.

OFFICIAL CCRF CLASSIFICATION
Supplemental Highly Recommended - 424 official CDEs New record Save record locally

Supplemental Highly Recommended: 424 official CDEs (89 clinician-completed and 335 participant-completed). LipoScope uses 445 REDCap dictionary rows to preserve the questions together with instructions, identifiers, calculated-field definitions and conditional logic. Conditional questions are included normally when their official branching conditions are met. Completion below is calculated only from applicable required response fields that can be assessed in this browser; it is not a percentage of all 424 CDEs.

OFFICIAL FORM **OFFICIAL SECTION** **FIND A QUESTION** **DISPLAY**

Diagnostics/Symptoms Diagnostics/Symptoms questions Search by question or variable REQUIRED ONLY MISSING OR INVALID ONLY

Diagnostics/Symptoms questions
50 field(s) currently shown from Diagnostics/Symptoms. Conditional questions appear only when their official branching conditions are met.

Clinician has confirmed that patient has a positive Lipedema diagnosis. Missing

Required

☒ Not entered ☐ Yes ☐ No

► Source details

What stage of Lipedema does the participant have currently? Please consider the legs/lower extremities when determining stage. Missing

Required

☒ Not entered ☐ Stage 1 ☐ Stage 2 ☐ Stage 3

☐ Unknown

Figure 19. Navigating CCRF forms and questions

8.3 Field controls and source details

Field type	Displayed behaviour
Radio	One coded choice, with a Not entered option.

Field type	Displayed behaviour
Dropdown	One coded selection from the official response list.
Checkbox	Multiple coded choices may be selected.
Text or number	Free entry with applicable validation, minimum, maximum or integer rules.
Notes	Multi-line text entry.
Date or date-time	Text entry with the displayed expected format.
Descriptive	Instruction or information only; no response is stored.
Identifier	Privacy message only; the value is not entered or stored in LipoScope.
File	Current CCRF Version 1 file/signature rows are identifier-marked and are displayed for source completeness but are not enterable or retained by LipoScope.
Calculated field	Displayed as unavailable unless the exact official calculation expression is supplied and validated.

Each field card can display Required or Optional status, current field status, notes, validation errors and a Source details section containing the official variable, field type, validation rules, minimum and maximum values, branching expression and coded choices. The View in CCRF data dictionary button opens the same variable in the dictionary module.

Matrix fields: Some official CCRF source fields belong to REDCap matrix groups. LipoScope preserves the corresponding matrix metadata in its embedded schema and CCRF Data Dictionary, but the Standardized Cohort Builder presents matrix members as separate response controls rather than reproducing the complete REDCap matrix. Shared matrix-level instructional text is not separately displayed in the Standardized Cohort Builder in v1.0.0. Users should therefore review the CCRF Data Dictionary or original CCRF source when the complete matrix instruction, recall period or administration context is required.

8.4 Official branching and inactive responses

Conditional fields appear only when their official branching condition is met. For example, a follow-up date may appear only after Yes is selected for the associated event. Conditional fields are normal CCRF questions and are not excluded for privacy.

If a response was entered and an earlier answer later changes so that the field is no longer applicable, LipoScope retains the response rather than deleting it silently. The response is marked as inactive with value, hidden from the currently applicable form and excluded from completion. The record summary and data-quality export identify it, and the user can clear it explicitly.

8.5 Record summary and saved-record ledger

The active-record summary reports applicable required-response completion, complete fields, missing fields and invalid fields. It also provides section-level summaries and warnings for stored responses in inactive branches. This percentage is a LipoScope workspace measure and is not an official CCRF score or diagnostic readiness score.

Ledger action	Function
Load	Makes the saved record the active record.
Codes + labels	Downloads a combined coded and readable file for that saved record.
Delete	Removes the saved record from the current browser profile.

Ledger action	Function
Download research package	Exports all eligible saved/current records in the complete package.
Clear saved records	Deletes all saved CCRF records from the current browser profile.
Source status	Shows whether a saved record matches the embedded schema checksum or has a version mismatch.

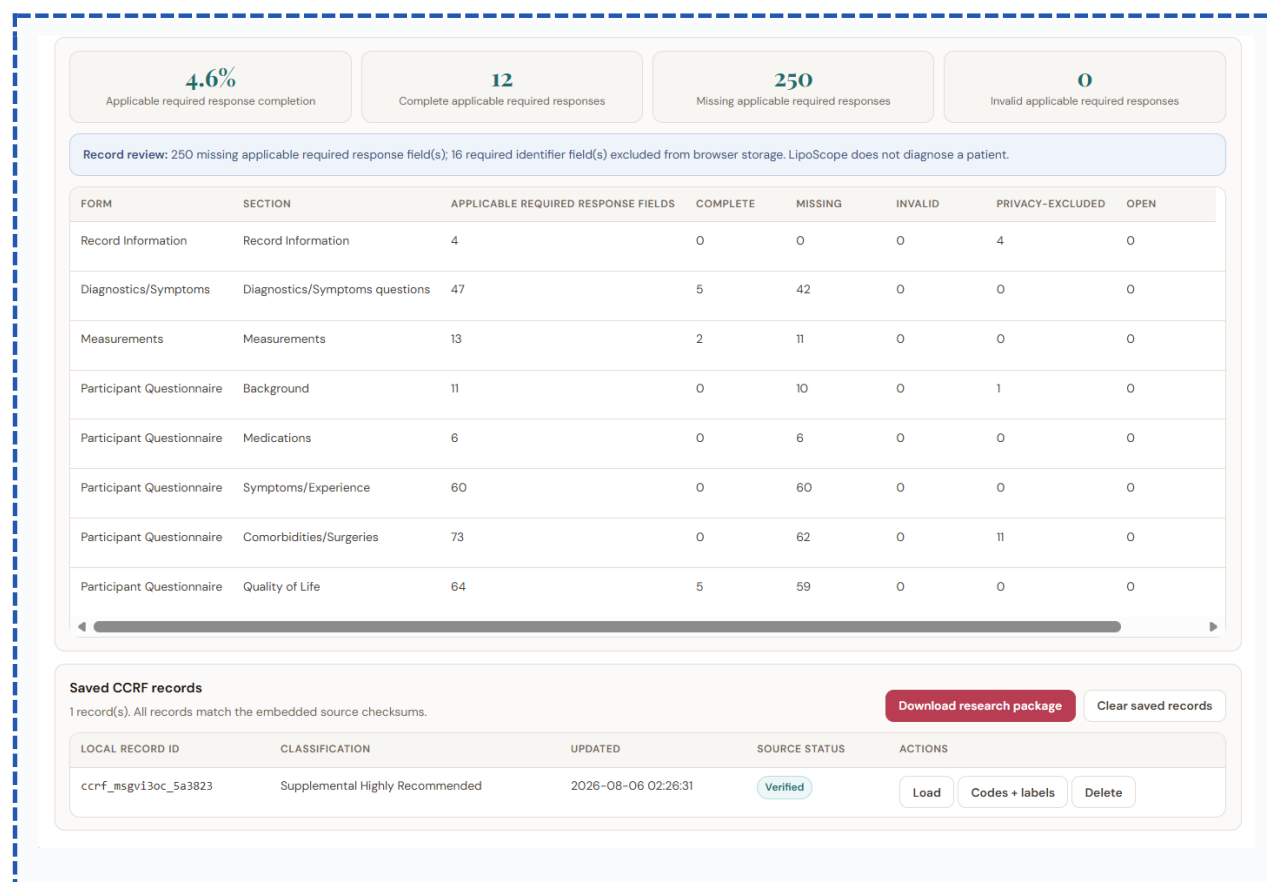


Figure 20. Active record summary and saved-record ledger

8.6 Cohort Design Gap Radar

The Gap Radar uses the same active record and source schema as the Builder. It reports issues by official form and section and can open the relevant Builder question or dictionary variable. It also summarises all saved records by classification.

Gap category	Meaning
Missing applicable required response	A required field is currently applicable but no valid response has been entered.
Invalid applicable required response	The value does not satisfy the source validation rule or allowed range.
Privacy-excluded identifier	The official field is an identifier and is intentionally unavailable for browser-local storage.

Gap category	Meaning
Stored response in inactive branch	A value is retained for a field that is no longer applicable under the official branch condition.
Version mismatch	A saved record was created with a different schema version or source checksum.

Select Export status CSV to download the Gap Radar status table. Use the Builder to correct issues and return to the Radar to confirm the updated state.

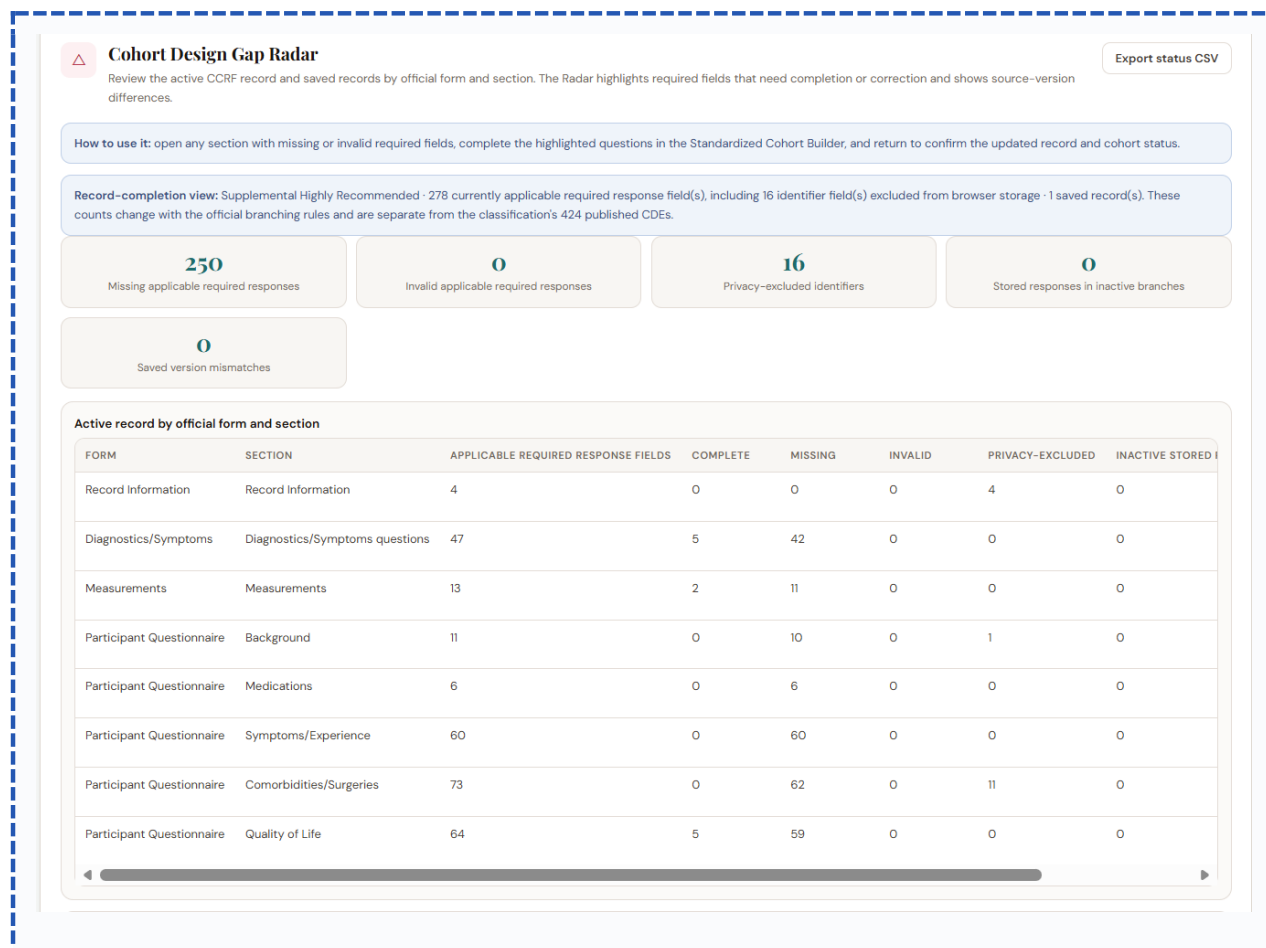


Figure 21. Cohort Design Gap Radar

8.7 CCRF Data Dictionary

The Data Dictionary displays the official source definitions for the selected classification. It can be reviewed independently of data entry or compared with the active record.

Dictionary control	Function
CCRF classification	Selects the source dictionary.
Official form	Limits rows to one form.
Official section	Limits rows to one section.

Dictionary control	Function
Field type	Filters text, notes, radio, dropdown, checkbox, slider, file, calc or descriptive rows.
Requirement or identifier	Shows all, required, optional or identifier rows.
Search dictionary	Searches variable, question wording or section.
Show active-record responses	Adds entered code, readable answer and field status when the active record uses the same classification.
Previous/Next	Moves through matching rows in the paginated table.

The table can show variable name, official question, form, section, field type, requirement, identifier status, source codes or calculation cell, validation and branching. When active responses are shown, the entered code, readable answer and current status appear beside the official definition.

CCRF Data Dictionary and Dataset Export
Understand how Standardized Cohort Builder responses are coded, compare the active record with the official CCRF dictionary, and download analysis-ready and human-readable research files.

How entered responses become research data Read before exporting

CCRF response codes provide consistent database values; readable labels explain what each code means. Both are generated from the same official field definition used by the Standardized Cohort Builder.

OFFICIAL VARIABLE	STORED CODE	READABLE ANSWER	DICTIONARY LINK
stage_current	2	Stage 2	2 = Stage 2

Recommended: complete research package

Downloads one ZIP containing coded data, readable data, paired codes and labels, long-format records, the official data dictionary, a data-quality report and a source manifest.

For statistical analysis

Use CCRF_records_coded.csv together with CCRF_data_dictionary.csv in R, Python, SPSS, Stata or another governed research environment.

For clinical or manual review

Use CCRF_records_codes_and_labels.csv or CCRF_records_human_readable.csv so every response can be interpreted without looking up numeric codes separately.

▼ Files included in the complete research package

- CCRF_records_codes_and_labels.csv - paired official codes and readable answers.
- CCRF_records_coded.csv - analysis-ready official codes.
- CCRF_records_human_readable.csv - readable answers.
- CCRF_records_long.csv - one field per row with code, label, status and source metadata.
- CCRF_data_dictionary.csv - variable definitions, codes, validation and branching.
- CCRF_data_quality_report.csv - missing, invalid, privacy-excluded, inactive-branch and version-mismatched entries.
- CCRF_source_manifest.txt - LipoScope version, CCRF version, classifications and source checksums.

Official CCRF classifications: Core, Supplemental Highly Recommended, Supplemental and Exploratory provide progressively more inclusive sets of data elements. The Lipedema Foundation recommends using the longest form feasible for the study design.

CCRF CLASSIFICATION	OFFICIAL FORM	OFFICIAL SECTION
Core	Diagnostics/Symptoms	All official sections

FIELD TYPE	REQUIREMENT OR IDENTIFIER	SEARCH DICTIONARY
All field types	All rows	Variable, question or section

☒ Show active-record responses, readable answers and field status

[Download complete research package -- recommended](#)
[Download codes and labels](#)
[Download analysis-ready coded data](#)
[Download human-readable data](#)

[Download official data dictionary](#)
[Download blank database columns](#)

Core: 288 official CDEs (86 clinician-completed; 202 participant-completed). The technical table contains 300 REDCap dictionary rows because it also preserves instructions, calculated fields, identifiers and conditional implementation rows. Browse variables, labels, response codes, validation and branching below.

Active record comparison: select the same CCRF classification in the Standardized Cohort Builder to display entered codes, readable answers and field status.

Figure 22. CCRF Data Dictionary controls and table

8.8 Dataset exports

LipoScope provides individual downloads and a recommended complete research package. Before export, the application warns when stored responses remain in inactive branches. If export continues, those values remain in the data but are marked in the data-quality file.

File	Purpose
CCRF_records_codes_and_labels.csv	Wide data with official codes and readable answers paired together.
CCRF_records_coded.csv	Analysis-ready wide data using official response codes.
CCRF_records_human_readable.csv	Wide data using readable response labels.
CCRF_records_long.csv	One field per row with record, variable, code, label, status and source metadata.
CCRF_data_dictionary.csv	Definitions, coded choices, validation, required status and branching.
CCRF_data_quality_report.csv	Missing, invalid, privacy-excluded, inactive-branch and version-mismatch entries.
CCRF_source_manifest.txt	LipoScope version, CCRF version, classifications and source checksums.

Export choice	Recommended use
Complete research package	Preferred archival and transfer package because it includes data, dictionary, quality report and source manifest.
Codes and labels	Manual review when both the stored value and its meaning are needed.
Analysis-ready coded data	Statistical analysis in R, Python, SPSS, Stata or another governed environment.
Human-readable data	Clinical or manual review without separate code lookup.
Official data dictionary	Database design, code interpretation and source review.
Blank database columns	Create an empty wide template using the selected classification variables.
Current record: codes + labels	Quick single-record download from the Builder.
Gap status CSV	Quality-control status from the Gap Radar.

Recommended: complete research package

Downloads one ZIP containing coded data, readable data, paired codes and labels, long-format records, the official data dictionary, a data-quality report and a source manifest.

For statistical analysis

Use CCRF_records_coded.csv together with CCRF_data_dictionary.csv in R, Python, SPSS, Stata or another governed research environment.

For clinical or manual review

Use CCRF_records_codes_and_labels.csv or CCRF_records_human_readable.csv so every response can be interpreted without looking up numeric codes separately.

▼ Files included in the complete research package

- CCRF_records_codes_and_labels.csv – paired official codes and readable answers.
- CCRF_records_coded.csv – analysis-ready official codes.
- CCRF_records_human_readable.csv – readable answers.
- CCRF_records_long.csv – one field per row with code, label status and source metadata.
- CCRF_data_dictionary.csv – variable definitions, codes, validation and branching.
- CCRF_data_quality_report.csv – missing, invalid, privacy-excluded, inactive-branch and version-mismatched entries.
- CCRF_source_manifest.txt – LipoScope version, CCRF version, classifications and source checksums.

Official CCRF classifications: Core, Supplemental Highly Recommended, Supplemental and Exploratory provide progressively more inclusive sets of data elements. The Lipedema Foundation recommends using the longest form feasible for the study design.

CCRF CLASSIFICATION	OFFICIAL FORM	OFFICIAL SECTION
Core	Diagnostics/Symptoms	All official sections

FIELD TYPE	REQUIREMENT OR IDENTIFIER	SEARCH DICTIONARY
All field types	All rows	Variable, question or section

☒ Show active-record responses, readable answers and field status

Download complete research package – recommended
Download codes and labels
Download analysis-ready coded data
Download human-readable data
Download official data dictionary
Download blank database columns

Core: 288 official CDEs (86 clinician-completed; 202 participant-completed). The technical table contains 300 REDCap dictionary rows because it also preserves instructions, calculated fields, identifiers and conditional implementation rows. Browse variables, labels, response codes, validation and branching below.

Active record comparison: select the same CCRF classification in the Standardized Cohort Builder to display entered codes, readable answers and field status.

Figure 23. CCRF export guide and download controls

8.9 Direct identifiers, files and calculated fields

Direct identifiers: Fields marked as identifiers (privacy fields) remain visible in the data dictionary for source completeness but cannot be populated in LipoScope. Exported CSV files can technically be edited outside LipoScope.

Files and signatures: Uploaded CCRF file content and electronic signatures are not retained in saved records or exported datasets.

Calculated fields: LipoScope does not invent or substitute formulas. When an official source dictionary marks a calculated field but does not provide an executable expression, the field remains unavailable.

9. About and workspace management

The About tab combines resource information, citation and licensing, topic definitions, user guidance, local profile controls, sharing and database management.

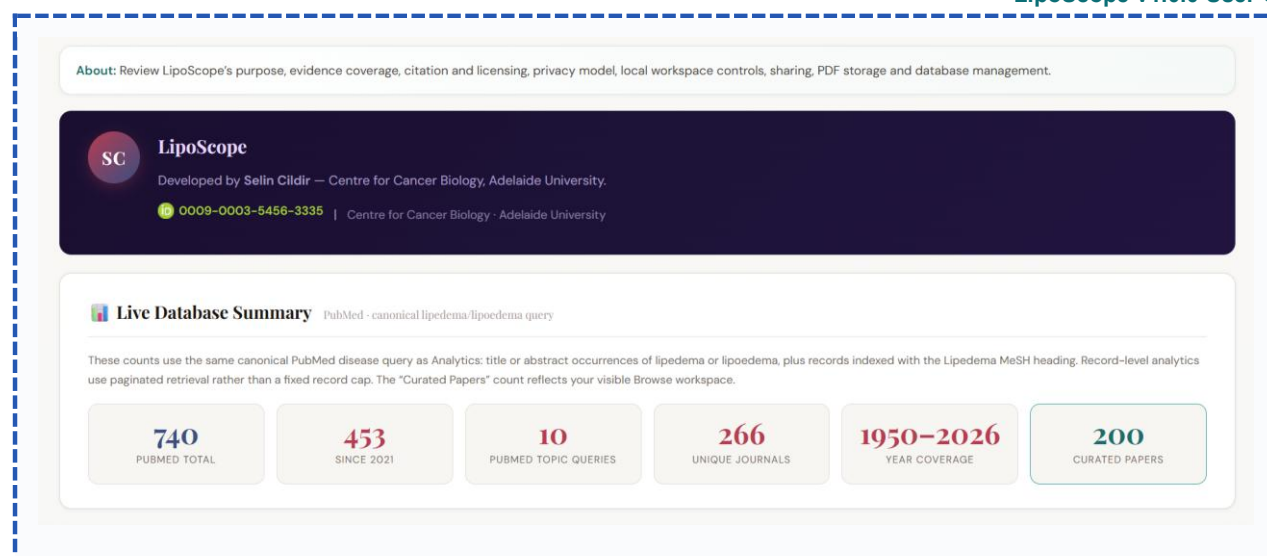


Figure 24. About tab overview

9.1 Citation, licensing and topic categories

The preferred software citation and repository link are provided in About. The software code is released under the MIT License, while original documentation is provided under CC BY 4.0. Official CCRF material, publications, abstracts, external figures and other third-party content retain their original terms.

The Research Topic Categories section displays the current topic-query counts and explains how topic classification is supported. Selecting a topic link opens the exact PubMed category query. Topic categories may overlap at the query level, while Browse records have no more than one displayed primary topic.

9.2 Public mode and optional local profile

LipoScope opens in Public Local Mode with no registration. Select Optional Local Label to create a named browser-local workspace. Re-entering the same label in the same browser can reopen its stored collection. The label is not an account, authentication method or central backup.

Control	Function
Optional Local Label	Creates or changes the active local browser label.
Default Website View	Returns to the default public workspace. Data under a named local label remain stored and can be reopened later in the same browser.
Share Current Browse View	Creates a URL that recreates the current Browse Papers structure for another user.
Open Shared Browse Link	Prompts for a LipoScope shared link and opens its Browse snapshot.
Shared-view banner actions	Keep the shared view, save it under a local label, return to the previous workspace or return to the default view.

What a shared Browse link contains: The link reproduces Browse Papers structure, including relevant visible, hidden and user-added paper information. It does not include CCRF records or local PDF files.

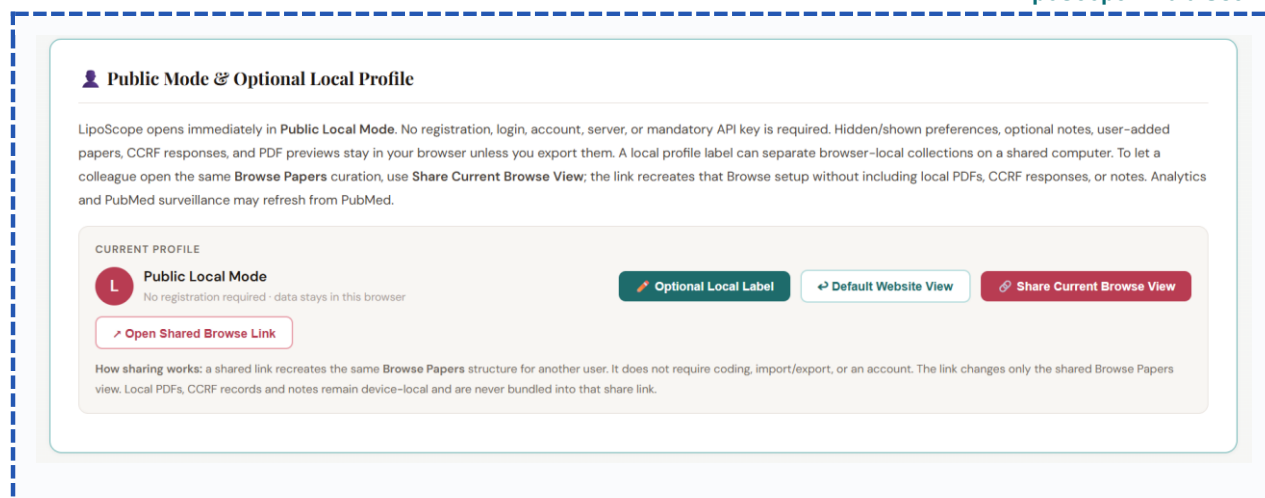


Figure 25. Local profile and Browse sharing

9.3 Database Management

Database Management displays the current numbers of visible papers, user-added papers, hidden papers and local PDFs. The controls change the current browser workspace only; PubMed-wide Analytics continues to use the broader disease-query layer.

Management control	Effect
Restore All Hidden Papers	Restores all hidden distributed papers.
Remove All User-Added Papers	Deletes all locally added paper records from the current workspace.
Hide All Default Literature	Creates a clean Browse workspace containing only user-added records.
Update Default Literature	Re-queries PubMed and rebuilds the browser-local PubMed-expanded portion of the default literature collection while retaining the fixed built-in seed and user-added papers. Because the expanded default set is replaced with current PubMed-derived records, the default papers visible in Browse may change. Using this control also re-enables default-literature updating if automatic refresh was previously paused by hiding all default literature.
Restore 200 Default Corpus	Reinstates the distributed starting library.
Hidden Papers list	Restores selected papers individually.
Keep uploaded PDF previews on this device	Enables persistent browser-local PDF storage for the current workspace.
Clear Stored PDFs For This Workspace	Deletes all persistent local figure PDFs associated with the current workspace.
Open PDF / Remove	Opens or deletes one stored PDF from the local library.

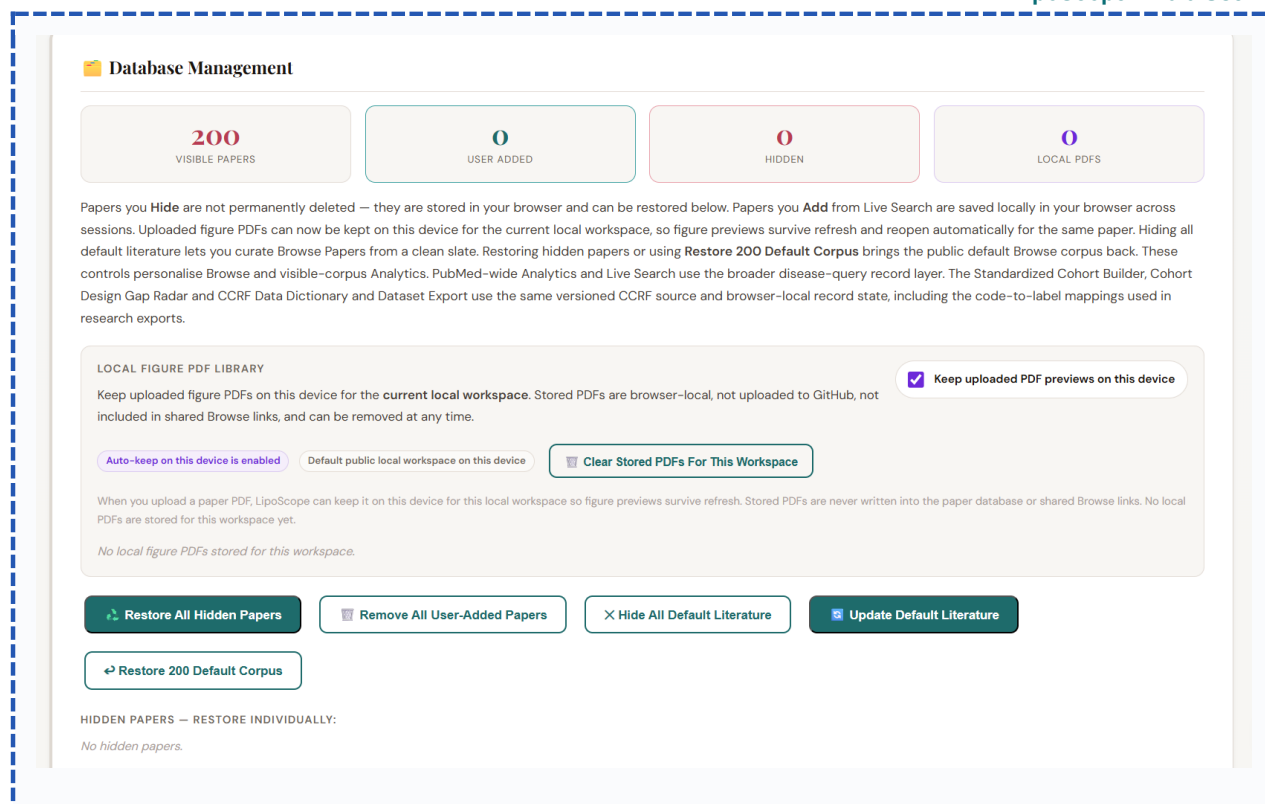


Figure 26. Database Management and local PDF library

10. Recommended workflows

10.1 Build a project-specific literature collection

- 1 Decide whether to use the starting library.** Keep it for orientation or hide all default literature in About for a clean project workspace.
- 2 Search Browse Papers.** Use keywords and filters to identify relevant papers already visible.
- 3 Add missing papers.** Use PubMed Live Search or New Paper Alerts.
- 4 Review article cards.** Open aims, methods, findings, source links and figures.
- 5 Remove articles.** Hide distributed papers or remove user-added papers that are outside the project or user's literature discovery scope.

6

Compare with the field. Use Analytics to see how the collection differs from the live PubMed landscape.

7

Share deliberately. Create a Browse link for curated literature and share it with collaborators, understanding that local PDFs and CCRF records remain local.

10.2 Use LipoScope while writing a manuscript or review

- **Locate references quickly:** Search by title, author, journal or findings.
- **Group reading by topic or design:** Use topic and study-type filters.
- **Check source provenance:** Open PubMed and DOI links from the paper window.
- **Inspect figures:** Use open-access figure lookup or local PDF preview where lawful.
- **Avoid relying on summaries alone:** Read and cite the original paper before making scientific claims.
- **Record the live query date:** PubMed counts and publication trends may change after manuscript submission.

10.3 Prepare a standardised research dataset

1

Confirm governance. Obtain the required ethics, consent, data-management and institutional approvals before collecting research data.

2

Choose the classification. Use the CCRF level specified by the study protocol.

3

Use a study ID. Do not enter names, contact details or other direct identifiers.

4

Complete applicable questions. Follow official form and section navigation; conditional questions appear automatically.

5

Review the record summary. Correct missing and invalid required fields.

6

Use the Gap Radar. Review section and cohort-level issues, inactive responses and source-version status.

7

Save and export. Save locally, then download the complete research package for governed storage and downstream work.

8

Retain source files. Keep the data dictionary, quality report and source manifest with the exported data.

10.4 Suggested audience-specific uses

Audience	Suggested workflow
Clinician or nurse	Use Browse Papers for targeted literature, consult readable CCRF definitions and contribute non-identifying research observations under an approved protocol.
Researcher	Curate a focused literature set, use live Analytics for field context and export coded CCRF data with its dictionary and quality report.
Patient organisation	Monitor new publications, organise public evidence and share Browse links without collecting participant identifiers.
Database manager	Review source variables, validation and branching; use exported dictionary and manifest when planning an institutional implementation.

11. Privacy, governance and safe use

LipoScope's local architecture gives users direct control over the current browser workspace, but local storage is not equivalent to secure institutional infrastructure. The following practices are recommended.

- **Do not enter direct identifiers:** Use coded study IDs in LipoScope and retain identifiable information only in the approved institutional system.
- **Use approved devices and browsers:** Avoid shared or public computers for research records unless local profiles and device access are appropriately managed.
- **Export and back up:** Move important research exports into governed storage. Browser data can be lost.
- **Review inactive responses:** Clear values that are no longer required or retain them with awareness that they will be marked in the data-quality file.
- **Check source versions:** Investigate any saved-record version mismatch before combining data.
- **Apply critical appraisal:** Article summaries, topic labels and study-type categories support navigation; they do not replace evaluation of the original methods and findings.

Data loss risk: Clearing site data, changing browser, using private mode, changing device or deleting a local profile may make stored data unavailable. The application does not provide central recovery.

12. Troubleshooting

Problem	Likely explanation and action
Live PubMed search or Analytics does not load	Check internet access, retry later, allow external requests, use the hosted application or try another current browser. LipoScope does not replace unavailable complete counts with estimates.
Browse shows no papers	Select Clear filters. If the starting library was hidden, open About and select Restore 200 Default Corpus or Restore All Hidden Papers.
A paper is missing from Browse	Search Live Search by title, PMID, author or keyword and add the record. Also check the Hidden Papers list.
Duplicate user-added papers are visible	Use Remove Duplicates in Auto-Discovery or remove the extra user-added card.
No figures are displayed	The paper may not have an identity-matched open-access PMC record. Open the source paper or upload a local PDF of the article for preview.

Problem	Likely explanation and action
Local PDF preview disappears after refresh	Enable Keep uploaded PDF previews on this device in About. Private mode or browser storage restrictions may still prevent persistence.
A shared Browse link is very long	Non-PubMed added records require more metadata in the URL. The link may still work; PubMed-based collections are usually shorter.
A shared link changed the visible workspace	Use the shared-view banner to save it under a local label, return to the previous workspace or return to the default public view.
A CCRF question is not visible	Its official branching condition may not be met, or the current form, section or display filters may exclude it.
A CCRF response is marked invalid	Open Source details to review validation, minimum, maximum or allowed choices, then correct the value.
A calculated CCRF field is unavailable	The embedded official source does not provide a validated executable expression. LipoScope intentionally does not infer a formula.
A record is missing after reopening	Confirm the same browser and local profile label are active. Browser storage may have been cleared.
ZIP export is unavailable	Use the individual CSV download buttons. The ZIP component may be blocked or unavailable in the current browser session.
Downloaded CSV opens incorrectly in spreadsheet software	Import the file as UTF-8 and preserve codes as text where leading zeros or exact values matter.
Saved record shows a version mismatch	Do not combine it silently. Review the source checksum and migrate the record under a documented process.

Reporting a software issue: Record the LipoScope version, browser, operating system, steps that caused the issue and any visible error message. Do not include identifiable participant data in screenshots or issue reports.

13. Glossary

Term	Meaning
Article card	A Browse Papers summary containing citation details, classification, a structured summary and source actions.
Browser-local	Stored in the current browser rather than automatically uploaded to a central LipoScope server.
CCRF	Common Case Report Form; the Lipedema Foundation Version 1 standard implemented in the CCRF Workspace.
CDE	Common data element; a standardised research variable or question.
Coded data	Data stored using official response codes, such as 2 rather than Stage 2.
Readable data	Data in which response codes are converted to their human-readable labels.
Long-format data	One variable-response observation per row rather than one record per row.
Branching logic	An official condition that determines whether a follow-up field is applicable.
Inactive with value	A retained response for a field whose official branching condition is no longer met.
Applicable required field	A required question whose branch is active for the current record.

Term	Meaning
Locally assessable	A field that LipoScope can validly collect and check in the browser; direct identifiers are excluded.
Gap Radar	The CCRF quality-review module for missing, invalid, excluded, inactive and version-mismatch states.
PMID	PubMed identifier for a publication record.
PMCID	PubMed Central identifier for an archived full-text record.
DOI	Digital Object Identifier used to resolve a publication.
Primary topic	The single displayed topic assigned to a paper when evidence is sufficiently distinct.
Unsupported record	A record that cannot be assigned a supported primary topic from available evidence.
Unclassified / Other	A record whose study type cannot be resolved confidently.
Visible corpus	The papers currently visible in Browse Papers after adding, hiding and filtering.
Default corpus	The distributed 200-paper starting library.
User-added paper	A PubMed record imported into the current browser through Live Search, Auto-Discovery or alerts.
Source manifest	A text file recording software version, CCRF version, classifications and source checksums.
Local profile label	An optional browser-only label used to separate workspaces; not a login or security account.
Shared Browse link	A URL that recreates a Browse Papers configuration without transferring CCRF records or local PDFs.

Appendix A. Quick guide for users

Task	Shortest route
Find a paper already in the workspace	Browse Papers → Search or filters → Open card.
Add a PubMed paper	Live Search → PubMed → Search → + Add to Browse.
Check newly published records	Notification bell → Check Now → Add to Browse.
Build a clean personal article collection	About → Hide All Default Literature → add papers through Live Search.
Restore the starting library	About → Restore 200 Default Corpus.
Compare collection with PubMed	Analytics → compare upper live charts with lower visible-corpus charts.
Create a research record	CCRF Workspace → choose classification → New record → enter applicable fields.
Find missing required answers	CCRF Workspace → Cohort Design Gap Radar.
Check an official variable	CCRF Data Dictionary → search by variable or question.
Export complete CCRF data	CCRF Data Dictionary and Dataset Export → Download complete research package.
Share literature curation	About → Share Current Browse View.

Task	Shortest route
Remove stored local PDFs	About → Clear Stored PDFs For This Workspace.

End of LipoScope v1.0.0 User Guide

Supplementary Methods S1

Application orientation and local workspace

Start Here orientation

The Start Here tab provides a simplified introduction to the application and explains how its literature and research-data functions are connected. It includes web links to clinical and community resources and a guided walkthrough covering five principal tasks: reviewing articles, adding PubMed records, interpreting literature analytics, preparing a Common Case Report Form (CCRF) research record and managing the local workspace. The walkthrough is implemented as static application content. The guidance distinguishes the preconfigured 200-paper starting (default) library from the broader PubMed literature landscape. It also explains that the default library can be hidden or customized when a user wishes to construct a project-specific article collection. Hidden papers can subsequently be restored individually or as a complete collection.

About tab and local workspace management

The “About” tab provides information on the application scope, privacy model, citation, licensing and local workspace state. It also contains controls for managing the default literature collection, user-added papers, hidden records and locally stored PDFs. Users can hide the complete 200-paper starting library, restore the full starting library, restore individual records or clear browser-local additions. These actions affect only the current browser profile and do not alter the distributed LipoScope application.

Literature curation, discovery and organization

Browse Papers workspace

The Browse Papers tab provides a project-specific workspace for literature curation and retrieval. The distributed version (v1.0.0) contains a bundled 200-publication default collection comprising a fixed built-in seed and a PubMed-expanded default layer. The bundled collection provides an immediate starting state for the article-card interface; however, the expanded layer is browser-local and refreshable rather than immutable. Shortly after application initialization, LipoScope checks the browser-local default-corpus state and when the previous successful or attempted check is older than seven days and automatic default refresh has not been disabled for that profile,

attempts to re-query PubMed for recent lipedema/lipoedema records and rebuild the expanded default layer while retaining the fixed built-in seed and separately stored user-added papers. Users may also trigger the same process explicitly using 'Update Default Literature'. Because this process reconstructs the expanded default layer rather than merely refreshing metadata for an unchanged set of records, the membership of the default collection displayed in Browse Papers may change over time and between browser profiles according to refresh history, internet availability and the PubMed results returned at the time of update. The default collection is intended as an orientation and demonstration set rather than as a systematic-literature corpus or a fixed denominator for literature analysis.

The starting library is operationally independent of the PubMed-wide literature retrieval used in the "Analytics" tab. Users may retain the complete starting library, hide individual records or hide all 200 publications through the "About" tab and construct an independent project-specific collection using records identified through "Live Search" or "New Paper Alerts". Records added by users are stored separately in the browser and are combined with the currently visible starting-library records only when the Browse Papers workspace is rendered. Hiding a distributed record removes it from the user's visible workspace without deleting it from the application and hidden records can subsequently be restored.

The visible literature collection can be searched by title, author, journal, summary and findings and can filter records by research topic, study type and publication year. These operations modify the current literature view without changing the underlying distributed or user-added records. The workspace is therefore intended to support literature curation, rapid article retrieval, manuscript preparation and the organization of publications relevant to a particular research domain.

Each publication is represented as a structured article card containing the available bibliographic and classification metadata. Opening a card presents the available study aims, methods, key findings information and links to PubMed, the DOI or other source records. Structured abstract summaries are produced using deterministic section and sentence handling rules and are intended to assist literature navigation rather than replace critical appraisal of the original publication.

For publications with a verified PubMed Central identifier, LipoScope can retrieve open-access article XML and may identify structured figure elements. Full-text records are checked against the available PMID, DOI and title information before figures are associated with an article card. When an open-access full-text source is unavailable, users may use a downloaded PDF of a specific article for device-local figure preview. Local PDF contents are not uploaded or transmitted by LipoScope.

Live Search, Auto-Discovery and New Paper Alerts

The Live Search tab provides three distinct literature-discovery pathways: an interactive PubMed search, an external Google Scholar search handoff and a topic-based PubMed Auto-Discovery function [1, 2]. These pathways operate in the user's browser and require an internet connection.

Interactive PubMed search

For an interactive PubMed search, LipoScope constructs the following expression:

```
<user-entered query>  
AND  
(lipedema[tiab] OR lipoedema[tiab] OR lipedema[MeSH Terms])
```

LipoScope submits the search through the National Center for Biotechnology Information (NCBI) ESearch utility and requests a maximum of 20 PMIDs [3]. The interface reports the total number of PubMed matches but displays only the returned subset of up to 20 records. Citation details are subsequently obtained through ESummary [3]. Interactive result cards display available title, abbreviated author list, publication year, journal and links to PubMed, DOI-resolved content and Google Scholar. Abstracts, Medical Subject Headings (MeSH) and publication-type metadata are not retrieved for all displayed search results; they are requested only when a user selects "Add to Browse".

Interactive PubMed results are held temporarily in the page's JavaScript memory and are replaced when another search is run. They are not saved to the browser-local literature collection unless the user explicitly selects "Add to Browse". Before an interactive import, LipoScope checks the PMID against the currently visible Browse corpus. It then attempts to retrieve the PubMed XML abstract and associated

metadata through EFetch, applies the versioned primary-topic and study-type classifiers and stores the resulting record in the current browser profile [3]. Added records persist in browser local storage and subsequently appear in Browse and visible-corpus Analytics. The current implementation's interactive duplicate check is based on PMID, normalized title and the currently visible corpus; a hidden built-in record is therefore not necessarily detected at this stage.

After an interactive record is added, LipoScope attempts to construct aims, methods, findings through deterministic, rule-based extraction from abstract text (**See section: Deterministic structured-summary extraction**). Despite the internal function name "generateAISummary", this process does not call a generative language model. If the PubMed abstract produces only fallback content, the application may request an abstract from Europe PMC using the PMID [4]. If this remains insufficient, it attempts to perform a Semantic Scholar title search and use the first returned abstract [5]. If sufficiently informative source text remains unavailable, generic fallback statements may be retained.

Google Scholar handoff

When Google Scholar is selected, LipoScope opens an external search using:

(<user-entered query>) (lipedema OR lipoedema)

The search is then opened in a new browser tab. LipoScope does not retrieve, parse, classify, deduplicate or locally store the Google Scholar result set and Google Scholar results cannot be imported directly through the displayed LipoScope result-card mechanism. Consequently, this pathway is an external search handoff rather than an integrated Google Scholar data-retrieval pipeline.

Topic-based PubMed Auto-Discovery

The Auto-Discovery function is PubMed-only and searches user-selected subsets of the ten fixed LipoScope topic queries. Users can request 10, 20, 30 or 50 of the most recent results per selected topic and may optionally restrict searches to publications from 2020, 2022, 2023, 2024 or 2025 onward. Auto-Discovery does not retrieve every match for a topic when its total exceeds the selected per-topic maximum. It obtains

the most recent requested PMIDs and summaries and then attempts to retrieve the PubMed abstract for each candidate.

Before import, Auto-Discovery rejects a candidate when its title and retrieved abstract do not contain at least one of lipedema, lipoedema, lipedematous or lipoedematous. This additional text test means that a record returned solely through Lipedema MeSH indexing can be excluded if the disease is not named in its title or abstract. Conversely, when abstract retrieval fails silently and the title contains a recognized disease expression, the record can still pass the disease-text check with incomplete metadata.

Auto-Discovery automatically imports qualifying candidate records without requiring individual user prompt or confirmation. Candidates are checked against the current Browse corpus by PMID and a normalized-title comparison and PMIDs encountered in more than one selected topic are processed only once during that discovery run. The normalized title is constructed by lowercasing the title, removing non-alphanumeric characters and retaining its first 80 resulting characters. A final deduplication pass compares user-added papers against the complete starting (built-in or default) library and removes duplicate local records while retaining the built-in record. Imported records are classified using the existing versioned topic and study-type classifiers and are saved as browser-local user-added records.

Auto-Discovery's displayed counters require interpretation: "Total PubMed hits searched" is calculated by summing the PubMed-reported totals for the selected topic queries, even though only the selected maximum of recent records per topic is retrieved and examined. Because the topic queries can overlap, this value is neither a count of unique publications nor the number of records actually processed. The skipped counter combines duplicate candidates and records rejected by the disease-text test, although the completion display labels the counter as skipped duplicates. These counters are therefore not reported as unique-literature flow counts.

New Paper Alerts

New Paper Alerts are implemented independently of the Live Search tab through the in-application bell menu. After LipoScope loads, the browser attempts an alert check

151 if at least 12 hours have elapsed since that browser profile's previous recorded check.
152 No checks occur while the application is closed.

153 The alerts query is:

154 (lipedema[tiab] OR lipoedema[tiab])
155 AND
156 ("<start date>"[PDAT] : "3000"[PDAT])

157 This differs from the canonical Analytics query because it does not include the
158 Lipedema MeSH-heading arm. It is also based on PubMed publication date (PDAT),
159 not the date on which a citation was created, entered or indexed in PubMed. Therefore,
160 the alert system is not described as an exhaustive detector of all newly indexed
161 Lipedema records. For instance, a newly entered citation carrying an older publication
162 date or a record identifiable only by MeSH indexing may not be detected by the alert
163 query.

164 During a normal automatic check, the start date is derived from the previous recorded
165 check date. During the first check or whenever "Check Now" is selected, the system
166 searches a 30-day publication-date window. A maximum of 15 results is retrieved. If
167 more than 15 matching records occur within the searched interval, only the most
168 recent 15 are examined.

169 A returned alert is treated as "new-to-LipoScope" when its PMID is absent from the
170 currently visible Browse corpus. Because hidden built-in records are excluded from
171 that corpus, an existing but hidden built-in PMID may be presented as a new alert.
172 Notification PMIDs are also compared with previously stored notifications to prevent
173 the same PMID from being added repeatedly to the notification list. Alert items contain
174 citation-summary information only; the abstract, topic and study type are retrieved or
175 generated after the user selects "Add to Browse".

176 Notifications, unread state and the previous-check timestamp are stored separately
177 for each local LipoScope profile. Up to 50 notification items are persisted. Marking
178 notifications as "read" changes only their unread status, while clearing notifications

removes the stored items but does not reset the previous-check timestamp. Alerts are not automatically imported into Browse.

The alert timestamp is recorded before the PubMed request is attempted. Therefore, a failed automatic request can still advance the recorded check time and suppress another non-forced automatic attempt for approximately 12 hours. A user can nevertheless initiate another request using “Check Now”; this forced check searches the preceding 30-day publication-date interval.

Deterministic structured-summary extraction

After a PubMed record is added to Browse, LipoScope attempts to construct an aims statement, methods description and up to five findings. Despite the internal function name *generateAISummary*, this process does not call a generative AI language model. It applies deterministic text-cleaning, section-selection, sentence-filtering and keyword-scoring rules to available abstract text. The output is primarily selected or truncated source sentences rather than an abstractive summary.

First, the text is normalized by removing HTML tags, decoding selected HTML entities, standardizing whitespace and identifying labels such as Objective, Aim, Background, Introduction, Methods, Design, Materials and methods, Participants, Results, Findings, Outcomes and Conclusions. Structured PubMed abstract sections are used where available. Labelled sections embedded in otherwise unstructured abstract text or existing stored summary fields are also detected and merged but an existing section is not overwritten by a subsequently detected section with the same label.

The text is divided into sentences using periods, question marks and exclamation marks while temporarily protecting decimal numbers, statistical decimals, initials and selected abbreviations. Fragments of 20 characters or fewer and fragments matching programmed malformed-text patterns are excluded.

The aims statement is selected preferentially from sections whose labels contain OBJECTIVE, AIM, PURPOSE, BACKGROUND or INTRODUCTION. Within this material, LipoScope prefers sentences containing expressions such as aim, objective, purpose, investigate, evaluate, assess, compare, determine, describe or characterize. Sentences dominated by selected participant, measurement or statistical expressions

209 such as “the study included,” “patients were,” “body mass index,” “odds ratio,”
210 “confidence interval” or “follow-up period” are excluded from these preferred aims
211 selection. If no preferred sentence is found, the first otherwise usable sentence is
212 selected. The aims output is limited to one sentence and approximately 380 characters
213 with word-boundary truncation when necessary.

214 Methods text is selected preferentially from sections whose labels contain METHOD,
215 DESIGN, MATERIAL or PARTICIPANT. A sentence is treated as method-like when it
216 contains both:

- 217 1. an action expression, such as conducted, performed, included, enrolled,
218 recruited, evaluated, assessed, measured, randomized, followed up,
219 compared, analysed or collected; and
- 220 2. a design, participant, data or measurement expression, such as study, trial,
221 cohort, case-control, cross-sectional, retrospective, prospective, survey,
222 participants, patients, controls, samples, data, MRI, ultrasound, sequencing or
223 GWAS.

224 Sentences containing explicit review-method terminology, including systematic
225 review, scoping review, meta-analysis, bibliometric analysis, literature search,
226 database search and PRISMA can also qualify. A further rule accepts sentences
227 combining a recognized design expression, such as single-centre, multicentre, before-
228 after, open-label, double-blind, prospective, retrospective, cross-sectional, case-
229 control, cohort or qualitative, with a study, analysis, participant or control expression.

230 Sentences beginning with result, conclusion, aim or background labels and selected
231 contextual phrases are rejected as methods. When methods must be extracted from
232 a complete unstructured abstract, up to the first three qualifying sentences are
233 retained, initially limited to approximately 560 characters. Structured methods material
234 may subsequently retain up to four usable sentences within approximately 620
235 characters.

236 Findings are sought first in sections whose labels contain RESULT, FINDING,
237 OUTCOME or CONCLUSION. If these sections yield no candidate, the complete

238 abstract is examined. Sentences shorter than 32 characters or longer than 560
239 characters are rejected, as are aims, background, methods and malformed fragments.

240 Remaining sentences receive additive heuristic scores:

- 241 • one point for terms such as results, findings or conclusions;
- 242 • two points when the sentence contains a numeral;
- 243 • three points for result-associated expressions such as odds ratio, confidence
244 interval, a *p*-value, prevalence, %, significant, higher, lower, increased,
245 decreased, associated, correlated, identified, found, sensitivity, specificity,
246 accuracy, follow-up or quality of life;
- 247 • one point for interpretive expressions such as conclude, suggest, indicate,
248 support or highlight.

249 Only sentences receiving a score greater than zero are eligible. The five highest-
250 scoring sentences are retained and then returned in their original source order. These
251 rules rank sentences that resemble findings; they do not independently verify that each
252 selected statement represents the publication's principal result.

253 The short card summary is derived from the extracted aims statement when a non-
254 fallback aim is available; otherwise, it uses the first extracted finding or the fallback
255 aims text. It is truncated to a maximum of approximately 220 characters.

256 When adequate content cannot be extracted, LipoScope uses explicit cautionary
257 fallback text. The fallback aims statement reports that the paper is indexed as a
258 lipedema/lipoedema publication and describes a broad focus inferred from
259 programmed title/topic patterns. The fallback methods statement reports that detailed
260 methods or, for a review-like record, a formal review/search method were not available
261 from the indexed abstract and directs the user to the source article. The fallback finding
262 states that key findings were unavailable from the indexed abstract and that the source
263 article should be used for investigation. These are disclosure statements and not
264 extracted study results.

265 If PubMed text produces only fallback content, LipoScope may attempt to request an
266 abstract from Europe PMC using the PMID and apply the same deterministic rules. If

that remains insufficient, it may attempt to search Semantic Scholar using up to the first 200 characters of the title, request one result and apply the rules to the first returned abstract. The Europe PMC request is PMID-based. The Semantic Scholar fallback does not perform an explicit PMID, DOI or exact-title identity verification before accepting the returned abstract.

Literature metadata retrieval and classification

LipoScope retrieves public bibliographic information through the NCBI PubMed E-utilities [3]. Retrieved information may include PubMed identifiers, titles, authors, publication dates, journal metadata, abstracts, structured abstract sections, MeSH details, publication types, author keywords, indexed substances and indexing status. The study-type classifier is coded to accept PubMed Central (PMC) JATS article-type and full-text section evidence when those fields are present in a browser-stored record [6]. However, the standard literature-retrieval pathway in the current v1.0.0 release does not itself populate these PMC-derived classifier fields. Routine classifications in this release are therefore based primarily on PubMed publication types, MeSH or indexing evidence, structured or unstructured abstract text, titles and other available PubMed metadata. The availability and completeness of PubMed metadata vary between records.

Study-type classification

Classifier inputs and candidate generation

LipoScope applies study-type classification through `classifyPaperTypeDetailed()`. Study-type evidence sources were assigned prespecified heuristic weights in LipoScope. The weights were used as ordinal decision priorities. Although the application interface describes a “classification hierarchy,” the function does not stop after finding the first available metadata source. It evaluates the available evidence sources, creates scored candidate classifications and then resolves or abstains from the competing candidates. The classifier first normalizes the supplied record into a common metadata structure containing:

- PubMed publication types
- MeSH descriptors

- 297 • structured abstract sections
- 298 • complete abstract text
- 299 • conditional JATS article type
- 300 • conditional full-text Methods, Results and Conclusion sections
- 301 • PubMed citation/indexing status
- 302 • metadata-completeness indicators

303 **PubMed publication-type evidence**

304 Publication-type and MeSH comparisons use case-insensitive exact label matching.
305 The classifier first adds candidates for specific NLM PubMed publication-type labels
306 (**Table 1**).

307 **Conditional PMC JATS article-type evidence**

308 If `jatsArticleType` is already present in the record, the following mappings are
309 evaluated. Research and review JATS types receive 62 points. Other recognized JATS
310 types receive 96 points (**Table 2**). PubMed retrieval pathway in the released
311 LipoScope version does not itself populate this field from PMC XML. This step
312 contributes only when a record already contains the relevant JATS value.

313 **MeSH study-design evidence**

314 The classifier next examines PubMed MeSH descriptors. The following descriptors
315 produce a GWAS/Cohort candidate with a score of 102:

- 316 • Genome-Wide Association Study
- 317 • Cohort Studies
- 318 • Longitudinal Studies
- 319 • Registries

320 The following descriptors produce an Original Research candidate with a score of 98:

- 321 • Cross-Sectional Studies
- 322 • Case-Control Studies
- 323 • Prospective Studies
- 324 • Retrospective Studies

- 325 • Comparative Study
- 326 • Multicenter Study
- 327 • Validation Study
- 328 • Evaluation Study

329 These comparisons use descriptor labels and do not require the descriptor to be
330 marked as a MeSH Major Topic.

331 **Abstract design evidence**

332 LipoScope normalizes abstract-section labels to uppercase and combines sections
333 whose labels contain:

- 334 • METHOD
- 335 • DESIGN
- 336 • MATERIAL

337 It separately collects Results or Findings sections for the operational-research test
338 described below.

339 The Methods/Design text is searched for explicit expressions in this fixed order:

- 340 1. Systematic Review
- 341 2. Consensus
- 342 3. Guidelines
- 343 4. Case Report/Series
- 344 5. Narrative Review
- 345 6. Study Protocol
- 346 7. Conference Abstract
- 347 8. Correction/Retraction Notice
- 348 9. Editorial/Commentary
- 349 10. Letter/Correspondence
- 350 11. Methods/Technical Article
- 351 12. Roadmap/Report
- 352 13. GWAS/Cohort
- 353 14. Original Research

354 The function returns the first matching type from that ordered text-rule sequence.

355 Examples of the programmed expressions include:

- 356 • systematic review, meta-analysis, PRISMA
- 357 • consensus statement, expert consensus, Delphi
- 358 • clinical practice guideline
- 359 • case report, case series
- 360 • narrative review, scoping review, literature review
- 361 • study protocol, trial protocol
- 362 • conference abstract, meeting abstract
- 363 • editorial, commentary, perspective
- 364 • letter to the editor, correspondence
- 365 • technical note, methods article
- 366 • white paper, working group report, position paper
- 367 • genome-wide association study, prospective cohort, registry-based study
- 368 • randomized controlled trial, cross-sectional study, case-control study,
- 369 observational study, pilot study and validation study

370 A structured-abstract result produces score 88 for GWAS/Cohort, score 82 for every
371 other identified type. If the structured Methods/Design sections do not yield a type, the
372 classifier examines the complete abstract. Before doing so, it removes sentences that
373 appear only to recommend future research, including expressions such as:

- 374 • future or further research
- 375 • additional studies
- 376 • studies are needed
- 377 • should be investigated, evaluated, studied, validated or examined

378 The remaining abstract is processed using the same ordered design-expression rules.

379 A record can also be classified as Original Research when both of the following are
380 found: (a) Methods text contains an operational research expression, such as
381 participants, patients, subjects, samples, cohort, collected data, measurements,
382 interventions, assays, sequencing, imaging, questionnaire or survey. (b) Results text

contains an observed-result expression, such as results, findings, observed, identified, detected, increased, decreased, associated, significant, confidence interval, odds ratio, hazard ratio or a p-value expression. Both conditions must be satisfied for this particular fallback rule.

Conditional stored full-text evidence

If full-text Methods or Results fields are already present in the record, LipoScope applies the same design-text rules to them. A full-text result produces: score 84 for GWAS/Cohort, score 78 for another identified type. As with JATS article type, the classifier supports this evidence but the released PubMed parser does not populate these full-text classifier fields. This is therefore a conditional code path.

Title evidence

A title match is accepted only for the following categories:

- Study Protocol
- Editorial/Commentary
- Letter/Correspondence
- Conference Abstract
- Correction/Retraction Notice
- Methods/Technical Article
- Systematic Review
- Consensus
- Guidelines
- Case Report/Series
- Narrative Review
- Roadmap/Report

An accepted title candidate receives 68 points. Although the title-rule function can recognize cohort or primary-study terminology, the final title-evidence filter does not accept title-only GWAS/Cohort or Original Research candidates.

Study type candidate ranking and conflict handling

411 After all applicable sources have been examined, candidates are sorted by
412 descending score. The implementation does not sum scores from multiple sources.
413 For instance, a publication-type score and an abstract score supporting the same
414 category do not become a combined score. The highest individual candidate score
415 remains the decision score. Additional evidence supporting the leading type is retained
416 in the classification output for provenance.

417 LipoScope searches for any candidate belonging to a different type whose score is
418 equal to or no more than four points below the leading score. This margin was
419 implemented as an operational conflict threshold and should not be interpreted as
420 statistical calibration or external validation.

- 421 • If no such competing type exists, the leading type is assigned.
- 422 • If such a competitor exists, LipoScope abstains and assigns Unclassified/
423 Other.
- 424 • The unresolved reason is recorded as conflicting-study-type-evidence.

425 For example, a leading score of 118 and a different candidate scoring 114 would cause
426 abstention. A different candidate scoring 113 would not, because it is five points below
427 the leader.

428 **Confidence tiers and retained output**

429 When a type is accepted, confidence is determined only from the leading score:

- 430 • high: score ≥ 105
- 431 • moderate: score 85–104
- 432 • low: score < 85

433 It is important to note that these are programmed score bands. They are not calibrated
434 probabilities of classification. The output also retains:

- 435 • the selected type
- 436 • decision score
- 437 • supporting evidence and evidence sources
- 438 • other matched types

- whether the classifier abstained
- the unresolved reason
- the classifier version

Unclassified/Other reasons

If no candidate is produced, the classifier assigns Unclassified/Other. The unresolved reason is selected sequentially:

1. **awaiting-medline-indexing** when citation status is Publisher, In-Process or In-Data-Review
2. **no-abstract-or-article-type** when information is conflicting and there is no abstract, publication type or JATS article type
3. **generic-journal-article-only** when all publication-type metadata consist only of Journal Article
4. **metadata-retrieval-incomplete** when both publication-type and citation-status information are absent
5. **no-informative-study-design** for the remaining unresolved cases.

Primary research topic classification

The primary research-topic classifier evaluates available PubMed indexing metadata and record-level text using fixed labels, Boolean query terms and case-insensitive JavaScript regular expressions encoded in LipoScope. These expressions perform literal pattern matching. LipoScope assigns a maximum of one primary research topic to each dynamically classified record using classifier version *mesh-hybrid-multilabel-query-membership-2026.4*. The ten possible topics are listed below and their topic-specific controlled metadata are presented in **Table 3**:

- Genetics & Genomics
- Pathophysiology
- Hormones & Estrogen
- Diagnosis & Clinical
- Treatment & Surgery
- Lymphatics & Vasculature
- Adipose Biology

- Epidemiology & Registry
- Guidelines & Consensus
- Comorbidities & QoL

The identifiers shown in **Table 3** are retained as reference metadata in LipoScope. Record-level MeSH matching compares descriptor and qualifier labels.

Evidence sources and search terminology

The primary research-topic classifier considers each topic separately. For every topic, it checks the available MeSH qualifiers, MeSH descriptors, publication types, PMID-level topic-query membership, title, author keywords, indexed chemicals and supplementary concepts, abstract sections and conditionally available stored full-text sections. **Tables 4** and **4A-4C** summarize how each evidence source is evaluated.

LipoScope uses several types of search terms. Each has a different function:

- **PubMed query terms** (titleQueryTerms) are sent to PubMed as part of topic-specific searches. They help retrieve papers that may relate to a topic. The exact query terms are shown in **Table 3**, and the Boolean construction of each query is shown in **Table 5**.
- **Title patterns** (titlePatterns) are predefined expressions applied locally to the title of each retrieved paper. The title expressions are shown in **Table 6**.
- **Ordinary abstract patterns** (abstractPatterns) are predefined topic-related expressions applied locally to abstract and other text fields. In a structured abstract or stored full-text section, at least two different ordinary patterns are required when no strong pattern is present.
- **Strong abstract patterns** (strongAbstractPatterns) are a smaller predefined set of more explicit topic-related expressions. One strong-pattern match is sufficient for a structured section to provide topic evidence. “Strong” and “ordinary” are labels assigned in the source code of LipoScope. They describe how the programmed matching rules treat an expression. The ordinary and strong patterns are shown in **Table 7**.
- **Author keywords** are keywords included in the PubMed record. LipoScope tests them against the combined title and ordinary abstract patterns.

- **Indexed-substance evidence** consists of PubMed chemical labels and supplementary concepts. These are also tested against the combined title and ordinary abstract patterns.

PubMed research category - query construction and classification

LipoScope builds one predefined PubMed query for each topic using the terms in **Table 3**. Each query contains two or three Boolean arms, shown in **Table 5**. A PMID satisfies a topic query when it satisfies any available arm. LipoScope retrieves the PMIDs returned by each query and records each PMID's topic memberships. The total number of papers returned by a query is not used as an individual-paper classification score. Instead:

- membership in exactly one topic query produces score 90 and specificity 4 for that topic
- membership in more than one topic query produces score 54 and specificity 2 for each matched topic

Thus, query membership is one source of evidence and it is not the final classification.

Record-level topic evidence

After the topic-query memberships have been attached to a PMID, LipoScope evaluates the other available evidence for each of the ten topics. The classifier begins each topic with a score and specificity of zero.

First, LipoScope establishes whether disease evidence is present. Disease evidence exists when the record contains a MeSH descriptor labelled Lipedema, or when the whole word lipedema or lipoedema occurs in the title or abstract. Disease evidence does not produce a topic score by itself. It is required before a general topic-related MeSH descriptor or configured publication type can contribute evidence (**Table 4**).

A topic-specific qualifier attached to the Lipedema MeSH descriptor produces the strongest programmed evidence. A matching qualifier receives score 120 and specificity 6 when PubMed marks that qualifier as a Major Topic or score 112 and specificity 6 otherwise. When both the Lipedema descriptor and a matching topic-

related descriptor are marked as Major Topics, the topic receives score 104 and specificity 5. A matching major topic descriptor accompanied by other disease evidence receives score 78 and specificity 4, while a matching non-major descriptor accompanied by disease evidence receives score 60 and specificity 3. A configured publication type accompanied by disease evidence receives score 100 and specificity 5. In the primary-topic classifier, configured publication-type evidence applies only to the Guideline and Practice Guideline labels for Guidelines & Consensus (**Tables 3 and 4**).

“Major Topic” in this context is the flag supplied in PubMed MeSH metadata. It does not indicate or imply that the publication is scientifically more important or methodologically stronger.

Title, author-keyword and indexed-substance evidence

LipoScope next applies the predefined title expressions in **Table 6** directly to the publication title. Matching is case-insensitive and uses fixed regular expressions. Each configured expression is counted once when it matches, even if the same expression appears several times in the title.

The direct title score is calculated as:

$$\min(88, 68 + 6n)$$

where (n) is the number of distinct title expressions matched. One match therefore produces score 74, two produce 80, three produce 86 and four or more produce the capped score of 88. Direct title evidence has specificity 4 (**Tables 4 and 6**).

Author keywords are joined into one text string and tested against the combined title expressions in **Table 6** and ordinary abstract expressions in **Table 7**. The author-keyword score is:

$$\min(76, 58 + 4n) \text{ with specificity 3 (Table 4).}$$

PubMed chemical labels and supplementary concepts are also joined into one text string and tested against the same combined expression set. Indexed-substance evidence is scored as:

558 min(68, 50 + 4n) with specificity 3 (**Tables 4, 6 and 7**).

559 The distinction between ordinary and strong patterns is not applied in the author-
560 keyword or indexed-substance branches. In these branches, every matching
561 expression from the combined title and ordinary abstract lists is treated as a pattern
562 match.

563 **Structured abstract and stored full-text evidence**

564 LipoScope converts abstract-section labels to uppercase and combines sections
565 whose labels contain the programmed terms shown in **Table 4A**. The selected section
566 groups are:

- 567 • Objective, Aim, Purpose, Conclusion and Interpretation
- 568 • Results or Findings
- 569 • Methods, Design or Materials
- 570 • Background or Introduction

571 For each section and topic, LipoScope checks the ordinary and strong abstract
572 expressions listed in **Table 7**. A section supplies topic evidence when either:

- 573 • at least one strong pattern matches or
- 574 • no strong pattern matches but at least two distinct ordinary patterns match

575 When a strong pattern is present, eight points are added to the section's base score.
576 Multiple strong matches do not add more than eight points. When there is no strong
577 match, ordinary patterns add three points per match, capped at 12 additional points.
578 The section-specific base scores, specificity tiers and resulting scores are reported in
579 **Tables 4A and 4B**.

580 Each section source creates no more than one evidence entry for a topic. For example,
581 three matching expressions in the Results section affect the Results-source score but
582 do not create three separate Results evidence entries.

583 LipoScope applies the same rules to stored full-text Results, Conclusions and Methods
584 fields when those fields are already present in the record. This is a conditional code
585 path. The standard PubMed abstract-retrieval pathway in the released LipoScope

586 v1.0.0 file does not routinely populate these PMC full-text classifier fields (**Tables 4**
587 and **4A–4B**).

588 **Complete-abstract fallback**

589 For each topic, the complete abstract is examined as a fallback only when that topic
590 has received no evidence from any preceding source. Therefore, the fallback is not
591 used for a topic that already has evidence from MeSH metadata, publication type,
592 query membership, title, author keywords, indexed substances, a structured abstract
593 section or a stored full-text section.

594 In the fallback abstract:

- 595 • one or more strong patterns produce score 48 and specificity 1 or
- 596 • when no strong pattern is present, at least two ordinary patterns are required,
597 producing $\min(44, 30 + 5n)$ with specificity 1.

598 Before selected abstract sections and the fallback abstract are searched, LipoScope
599 removes sentences matching the future-recommendation expressions shown in **Table**
600 **4C**. These include expressions regarding future or further research, additional studies,
601 studies being needed and matters that should be investigated, evaluated, studied,
602 validated or examined (**Tables 4, 4C** and **7**).

603 **Combining evidence for each topic**

604 Whenever an evidence rule matches, LipoScope records an evidence entry for that
605 topic. The classifier does not add all original evidence-source scores together.
606 Instead, for each topic it retains:

- 607 • the highest individual source score
- 608 • the highest specificity tier and
- 609 • all evidence entries generated for provenance

610 The highest score and highest specificity can come from different evidence sources.

611 LipoScope then adds three corroboration points for each evidence entry after the
612 first. This bonus is capped at 12 points and the final topic score is capped at 125:

613 final topic score = $\min [125, \max(\text{individual source scores}) + \min(12, 3(\text{evidence entries} - 1))]$

614 The practical effect of the corroboration bonus is to give limited additional weight
615 when more than one item of evidence supports the same topic. For example,
616 agreement between a MeSH descriptor, a title expression and an abstract-section
617 expression strengthens that topic's candidate score. The cap prevents the score
618 from increasing without limit.

619 The bonus is based on evidence entries rather than repeated expressions. For
620 example, three matching title expressions generate one title evidence entry with a
621 title score of 86; they do not generate three corroboration entries. The corroboration
622 values are prespecified heuristic rules and are not empirically calibrated (**Table 8**).

623 **Creating and comparing topic candidates**

624 A topic becomes a candidate whenever at least one evidence entry has been
625 generated for that topic. Candidates are ordered sequentially by:

- 626 1. descending specificity
- 627 2. descending final topic score and
- 628 3. descending fixed topic-priority value

629 The fixed topic-priority order is shown in **Table 9A**. Priority is used only when
630 specificity and score are equal. It is not added to the evidence score and does not
631 indicate scientific importance, research priority or classification accuracy. Because
632 specificity is considered before score, a topic with greater specificity can be placed
633 ahead of another topic that has a higher score but lower specificity (**Tables 9** and **9A**).

634 **Selecting the primary topic**

635 The leading candidate must have a final score of at least 58. If this minimum is met,
636 the leading candidate is accepted when at least one of the following conditions is
637 satisfied:

- 638 • no second candidate exists
- 639 • the leading candidate has greater specificity than the second candidate

- the leading candidate's score exceeds the second candidate's score by at least eight points
- the leading candidate has unique topic-query-membership evidence or
- the leading candidate has specificity of at least 5 and exceeds the second candidate by at least three points

If the leading candidate does not have a score of at least 58 or if none of the separation conditions is met, LipoScope does not force the record into a topic. It assigns **Unsupported records** with the reason *multiple-equivalent-topic-evidence* when competing candidates cannot be adequately separated (**Table 9**).

Confidence and retained classification information

When a primary topic is accepted, LipoScope assigns a confidence label according to the final topic score:

- high: score ≥ 90
- moderate: score 68–89
- low: score below 68

These labels are fixed score bands. They are not calibrated probabilities of classification.

The stored output includes the accepted primary topic, final score, supporting evidence, evidence available for other topics, secondary topics, confidence band, resolution status and classifier version (**Table 10**).

For an accepted classification, non-primary candidate topics with scores of at least 45 are retained as secondary topics. Secondary topics record additional topic evidence but do not contribute additional counts to the mutually exclusive primary-topic distribution.

When the classifier abstains because the candidates cannot be separated, all generated candidate names are retained in the detailed *secondaryTopics* field without applying the usual score-45 threshold. The primary result nevertheless remains **Unsupported records**.

Unsupported records

When no topic candidate is generated, LipoScope selects the unresolved reason in the following programmed order:

1. **awaiting-medline-indexing** when citation status is Publisher, In-Process or In-Data-Review
2. **abstract-unavailable** when both abstract text and author keywords are absent
3. **pubmed-not-medline** when citation status is PubMed-not-MEDLINE
4. **metadata-retrieval-incomplete** when MeSH headings, publication types and citation status are all absent or
5. **no-category-specific-evidence** for the remaining cases

When candidates exist but cannot be separated, the reason is multiple-equivalent-topic-evidence (**Table 11**). The order of these checks affects the reported reason. For example, a PubMed-not-MEDLINE record without an abstract or author keywords receives abstract-unavailable, because the absence-of-content condition is evaluated before the PubMed-not-MEDLINE condition.

Relationship between PubMed category queries and primary-topic classification

Overall, ten PubMed research-category queries and the primary-topic classifier produce different outputs:

- A topic-query total is the number reported independently by PubMed for that query.
- Topic-query result sets can overlap because one PMID can satisfy several queries.
- Query membership contributes one item of record-level classifier evidence.
- The aggregate query total is not used as a classification score.
- Primary-topic classification combines query membership with available metadata and textual evidence.
- Each processed PMID contributes once to one accepted primary topic or to Unsupported records.

- Secondary topics preserve non-primary evidence but do not contribute additional primary-topic counts.

Accordingly, topic-query totals should not be summed to estimate the number of unique publications and they should not be interpreted as equivalent to the mutually exclusive primary-topic distribution. The displayed query categories describe overlapping PubMed search results, whereas the primary-topic distribution describes the final record-level classifier output (**Tables 3-11**).

All topic scores, specificity tiers, corroboration values, priority values and acceptance thresholds are fixed heuristic rules encoded in the released LipoScope file. They are not probabilities, measures of scientific importance or externally validated estimates of classification accuracy.

Analytics landscape

The Analytics tab distinguishes between two independently calculated descriptive views:

1. the broader PubMed-derived lipedema literature landscape
2. the literature currently included in the user's visible Browse corpus

The two denominators are displayed separately and are not combined. LipoScope does not calculate a formal over-representation measure, inferential comparison or significance test between them. Any statement that the local collection appears over- or under-represented is therefore a descriptive interpretation made by the user.

Headline counts and denominators under Analytics tab

The headline **PubMed Total** begins with the count returned directly by PubMed ESearch. As an implementation rule, LipoScope displays the larger of:

“the PubMed ESearch count” and “the number of PubMed records available in the browser record store”

The recent-publication value similarly uses the larger of the date-restricted ESearch count and the recent-record count calculated from available browser records.

724 The recent interval is defined dynamically as:

725 *current calendar year – 5 through current calendar year [PDAT]*

726 Both endpoints are inclusive. For example, the interval in 2026 is 2021:2026[PDAT];
727 it is not a rolling 60-month interval.

728 Publication timeline, journal, primary-topic and study-type distributions are calculated
729 from the PubMed records available in the browser after current or previous successful
730 retrieval and processing. Consequently, the headline PubMed query count and the
731 record-level chart denominators are expected but are not programmatically required
732 to agree. A retrieval failure, incomplete metadata response or reuse of an earlier
733 record cache may temporarily produce a difference between them.

734 The displayed recent-publication value is not calculated by summing the
735 corresponding Publication Timeline bars. LipoScope reports the larger of: (i) the direct
736 PubMed ESearch count for the configured publication-date interval and (ii) the number
737 of browser-available PubMed records whose extracted year is at or after the cutoff
738 year. By contrast, the timeline is constructed only from browser-available records for
739 which a valid publication year has been resolved. The values may therefore differ
740 because not every PMID counted by PubMed has been retrieved and processed in
741 the browser, because some processed records lack a resolvable year or because the
742 direct count and cached record set were obtained at different times. The local recent-
743 record calculation also has no explicit upper-year bound, while timeline validation
744 accepts a year up to the current year plus one. Accordingly, the recent value and the
745 sum of the timeline bars should be treated as separate descriptive quantities rather
746 than expected to reconcile exactly.

747 The application calculates the browser-processed record count internally as
748 liveRecordCount, although the current Analytics interface does not display this value
749 beside the PubMed Total.

750 **Publication timeline**

751 Publication year is resolved from the PubMed XML publication year and, if required,
752 the ESummary pubdate, epubdate and sortpubdate fields. A year is treated as valid

when it falls between 1800 and one year beyond the current calendar year. Each record with a valid year contributes once to the corresponding annual count. Records without an identifiable valid year remain represented in an unidentified-year count but are excluded from the annual bars. Zero-count years are inserted between the earliest identified year and the current year. Timeline bar heights are scaled relative to the largest annual count.

Journal distribution

Journal information is obtained from the metadata available for each PubMed record. LipoScope first selects an available journal label in the following order: PubMed journal title, ESummary full journal name, stored journal label, MEDLINE title abbreviation, ISO abbreviation and ESummary source label. A fixed source-code alias lookup then replaces seven selected normalized title variants with designated display titles. This lookup is not a journal whitelist; journal titles absent from it remain eligible for processing.

Each record is assigned a journal-identity key using the first available source in this order:

1. the NLM unique journal identifier [7, 8]
2. linking ISSN [9, 10]
3. an available ESummary print ISSN or electronic ISSN
4. an existing stored journal key
5. a normalized journal title

Before an ISSN is used as a key, non-alphanumeric characters are removed and the value is converted to uppercase. Journal-title normalization converts the label to lowercase, replaces the word “and” with “&”, removes other non-alphanumeric characters and normalizes surrounding spacing.

When a publication matches the PubMed-wide LipoScope query and its journal information is successfully retrieved into the PubMed record set processed by Analytics, the journal contributes to the distribution. It appears in the Top Journals panel if its record count places it among the ten largest displayed journal groups when

782 Analytics is next successfully refreshed; this is not a continuous background update
783 while the application is closed.

784 The **Unique Journals** statistic counts the number of distinct journal-identity keys in
785 the materialized PubMed record set. For the Top Journals display, records in those
786 identity buckets are subsequently aggregated by their selected display label, ranked
787 by descending record count and limited to the ten most frequent labels. Consequently,
788 the Unique Journals statistic is identity-key based, whereas the displayed ranking is
789 label based.

790 For each displayed journal:

791 $\text{journal percentage} = (\text{records assigned to that harmonized journal}$
792 $\div \text{all browser-processed PubMed records with an identified journal}) \times 100$

793

794 Records for which no journal label can be identified are counted separately and
795 excluded from the percentage denominator. The length of each journal bar is scaled
796 relative to the largest count among the ten displayed journals. The printed percentage
797 instead uses all records with an identified journal in the materialized PubMed record
798 set as its denominator.

799 **Visible LipoScope corpus**

800 This corpus consists of:

- 801 • non-hidden built-in/default papers
- 802 • browser-local user-added papers.

803 The charts in visible LipoScope corpus do not use the complete PubMed result set.
804 They are recalculated when the underlying visible corpus changes through actions
805 such as adding, hiding, restoring or expanding papers.

806 The topic donut counts the stored primary-topic value of each visible Browse record.
807 Percentages are rounded to whole numbers. All topic values can contribute slices but
808 only the ten most frequent values are listed in the legend.

The Browse year chart includes records with a populated year. It displays only years represented in the visible corpus, does not insert zero-count intervening years and does not separately report records without a populated year.

PubMed disease query and record retrieval

The PubMed-derived view uses the fixed canonical disease query:

(lipedema[Title/Abstract] OR lipoedema[Title/Abstract] OR "Lipedema"[Mesh])

When a refresh is required, the user's browser attempts exhaustive paginated retrieval of the PMIDs reported for this query. ESearch requests retrieve identifiers in pages of up to 500. ESummary metadata are requested in batches of up to 120 records and PubMed XML metadata and abstracts are requested through EFetch in batches of up to 60 records. Requests are rate-spaced and retried up to three times when a request fails.

A returned PMID is converted into a browser-stored LipoScope record when its required summary information is available and the retrieved metadata satisfy an additional disease-specific check. This check requires either a Lipedema MeSH heading or a title/abstract containing lipedema, lipoedema, lipedematous or lipoedematous. The resulting browser record store may contain records retrieved during the current session and records retained from previous successful retrievals. It is not a centrally maintained LipoScope database.

Caching, automatic updating and failure states

LipoScope is a static browser application rather than a continuously running literature-monitoring service. If the application is not open, LipoScope does not perform background PubMed retrieval.

The following browser caches ordinarily have a 12-hour validity period:

- the PubMed-derived record store
- the calculated PubMed Analytics statistics
- exact topic-query counts and PMID-to-topic memberships

836 A degraded or non-live Analytics result is cached for 20 minutes.

837 Refresh attempts occur:

- 838 • approximately 1.5 seconds after application startup
- 839 • at 12-hour intervals while the page remains open
- 840 • when the page becomes visible
- 841 • when the browser window regains focus
- 842 • when network connectivity returns

843 These events do not necessarily force a new PubMed request when a relevant cache
844 remains valid. New PubMed records matching the existing canonical query and new
845 journals represented by those records can be incorporated automatically following an
846 eligible and successful refresh.

847 The Live interface state indicates that the live-query Analytics workflow completed. It
848 may include browser records or topic memberships reused from still-valid live-derived
849 caches and does not guarantee that every PMID and metadata field was re-
850 downloaded during the current page view.

851 If the overall live Analytics operation fails and a previously successful result is
852 available, LipoScope can display that result as Cached live. If no successful Analytics
853 cache is available, it can calculate an explicitly incomplete local result from browser-
854 stored PubMed records. In this state, complete PubMed totals and exact current
855 category-query counts are not guaranteed.

856 An additional implementation limitation is that an inner PubMed record-refresh failure
857 can return the previously stored record set. If subsequent headline count requests
858 succeed, the combined Analytics result may still enter the Live state even though the
859 record-level charts use an older record set. Accordingly, the Live badge should not be
860 treated as proof that all record-level denominators were refreshed simultaneously.

861 Existing records considered stably indexed are not routinely re-downloaded at every
862 12-hour refresh. New matching PMIDs are discovered automatically but later PubMed
863 corrections to the metadata of an already cached stable record may not be
864 incorporated until the browser store is rebuilt.

Overall, LipoScope Analytics outputs are descriptive, time-dependent and subject to changes in PubMed content and indexing. The application uses browser-based retrieval and caching rather than a continuously operating central database. Headline PubMed counts and record-level chart denominators may temporarily differ following incomplete retrieval or cache reuse. Topic-specific PubMed queries overlap and must not be summed. Primary-topic and study-type assignments are generated by versioned, rule-based classifiers whose behaviour is tested against selected programmed examples but whose ground-truth accuracy has not been externally validated. Comparisons between the PubMed-derived landscape and the visible Browse corpus are visual and descriptive.

Common Case Report Form (CCRF) Workspace and standardized cohort preparation

The Standardized Cohort Builder under the CCRF Workspace tab allows user to select one of the four official CCRF classifications or source dictionaries originally provided by the Lipedema Foundation and create a browser-local research record based on the corresponding schema [11, 12]. Each record is assigned a generated study record identifier, selected classification, CCRF version, schema version, source checksum, timestamps and a response object.

The Builder organizes questions according to the official forms and section headings. Users can navigate between sections, search by variable name or field label and restrict the interface to required or unresolved fields. The field count shown for a section is calculated from the controls that are currently rendered after branching and display filters have been applied.

Browser controls are generated from the embedded source schema but the visual browser widget does not necessarily reproduce the corresponding native REDCap control. In LipoScope v1.0.0, source fields typed as “slider” remain identified as sliders in the embedded CCRF schema and data dictionary but are presented in the Standardized Cohort Builder as numeric entry controls rather than graphical range sliders. Source minimum and maximum constraints are applied when they are supplied in the source dictionary; when such bounds are absent, LipoScope does not infer additional slider limits. Radio buttons and dropdowns use the official coded choices, checkbox fields preserve the source code–label pairs and numeric, date, text

and notes fields use appropriate browser inputs. Descriptive source rows are displayed as information rather than as response variables. Matrix metadata are retained in the embedded CCRF schema and data-dictionary outputs but LipoScope v1.0.0 does not reproduce the complete native REDCap matrix presentation in the Standardized Cohort Builder. Matrix members are presented as separate browser controls and matrix-level instructional text stored in the source matrix metadata is not separately rendered above those controls. Consequently, when a source matrix contains shared instructions, recall periods or other group-level administration context, users should consult the CCRF Data Dictionary or original CCRF source material when that complete context is required. Accordingly, source/schema preservation of matrix metadata should not be interpreted as exact reproduction of the participant-facing REDCap matrix interface.

Conditional CCRF fields and branching

Conditional CCRF questions are controlled using the official branching expressions supplied in the source dictionaries. A conditional question is displayed only when its branching expression evaluates as applicable against the current record responses. Conditional fields are not treated as privacy-restricted fields.

When a response has been entered and a subsequent change to an earlier answer makes that field inapplicable, LipoScope retains the previous value rather than silently deleting it. The retained value is marked as inactive, hidden from the currently applicable form and excluded from the completion denominator. It is also identified in the record summary and data-quality output. Users can explicitly clear an inactive stored response when it is no longer required.

Required-field assessment and Gap Radar for CCRF workspace

For every currently applicable input field, LipoScope assesses whether the field is required, whether a response is present and whether the response satisfies the source validation type and limits. The assessment also identifies fields excluded as direct identifiers and inactive conditional fields that retain a previous response.

The displayed completion percentage is calculated as the number of completed, currently applicable, required and locally assessable fields divided by the total number of currently applicable, required and locally assessable fields. This is a LipoScope

workspace-completion measure. Direct identifiers excluded from local storage and fields in inactive branches do not contribute to this denominator.

The Cohort Design Gap Radar uses the same field-level assessment as the Standardized Cohort Builder. For the active record, it reports missing required responses, invalid values, direct identifiers excluded from local collection and retained values in inactive branches. The results are grouped by official form and section and users can navigate from the Gap Radar to the relevant data-entry location. For saved records, the Gap Radar also identifies records created under a different schema version or source checksum.

Direct CCRF identifiers and calculated fields

Fields marked as direct identifiers in the official source dictionaries remain visible in the original CCRF Data Dictionaries but these are not populated in saved browser-local records or exported datasets provided by LipoScope. This includes privacy source fields such as names, contact information and dates of birth when they are marked as identifiers by the official dictionary. Signatures and uploaded file contents are not retained as saved CCRF responses.

LipoScope does not infer or substitute conventional formulas for official CCRF calculated fields. Across the four cumulative source dictionaries, these fields are therefore displayed as unavailable calculations and are not automatically populated.

CCRF record storage and exports

Before a CCRF record is saved, it is passed through a sanitization function that removes identifier values and unsupported file content. Sanitized records are stored in a profile-specific browser ledger. Users can create, load, replace or delete individual records and can clear the complete local CCRF ledger.

The CCRF Data Dictionary displays the selected classification's source variable names, forms, sections, field types, labels, response choices, validation settings, required status, identifier status, branching expressions and current response status. It can be searched and reviewed independently of the active data-entry form.

LipoScope can export coded wide-format data, human-readable wide-format data, combined code-and-label data, long-format records, a classification-specific data dictionary, a blank data template, a field- and section-level data-quality report, a Gap

Radar report and a source manifest containing LipoScope and the Lipedema Foundation CCRF version and checksum information. Before export, the application warns the user when stored responses remain in inactive branches. When the export proceeds, those values are retained and identified in the data-quality output. Direct-identifier values and uploaded file contents remain blank.

References

1. PubMed. (2025). *PubMed overview*. PubMed; National Library of Medicine.; Available from: <https://pubmed.ncbi.nlm.nih.gov/about/>
2. Google Scholar. Available from: <https://scholar.google.com/>.
3. Sayers, E. (2022, November 17). *A General Introduction to the E-utilities*. Nih.gov; National Center for Biotechnology Information (US). Available from: <https://www.ncbi.nlm.nih.gov/sites/books/NBK25497/>
4. Europe PMC. [cited 2026; Available from: <https://europepmc.org/>.
5. Semantic Scholar. [cited 2026; Available from: <https://www.semanticscholar.org/>.
6. *Journal Article Tag Suite*. [cited 2026; Available from: <https://jats.nlm.nih.gov/>.
7. *NLM Catalog*. [cited 2026; Available from: <https://ncbi.nlm.nih.gov/nlmcatalog/>.
8. *NLM Unique ID - PubMed XML specification*. [cited 2026; Available from: <https://dtd.nlm.nih.gov/ncbi/pubmed/doc/out/250101/index.html>.
9. *ISSN*. Available from: <https://www.issn.org/understanding-the-issn/assignment-rules/the-issn-l-for-publications-on-multiple-media/>.
10. *ISSN Linking - PubMed XML specification*. [cited 2026; Available from: <https://dtd.nlm.nih.gov/ncbi/pubmed/doc/out/250101/el-ISSNLinking.html>.
11. Galia, S., et al., *The lipedema common case report form as a research tool: standardizing lipedema data collection*. Front Glob Womens Health, 2026. **7**: p. 1833913.
12. *LF Common Case Report Form (CCRF)*. [cited 2026 1 June]; Available from: <https://www.lipedema.org/ccrf>.

Tables for Supplementary Methods S1

Table 1: Candidate creation from PubMed publication types

PubMed publication type	Candidate output	Score
Systematic Review, Meta-Analysis, Network Meta-Analysis	Systematic Review	120
Published Erratum, Retracted Publication, Retraction of Publication, Expression of Concern, Corrected and Republished Article	Correction/Retraction Notice	120
Consensus Statement, Consensus Development Conference, Consensus Development Conference, NIH	Consensus	118
Guideline, Practice Guideline	Guidelines	118
Case Reports	Case Report/Series	116
Editorial, Comment	Editorial/Commentary	116
Letter	Letter/Correspondence	116
Clinical Trial Protocol	Study Protocol	116
Meeting Abstract	Conference Abstract	112
Technical Report	Methods/Technical Article	112
Randomized Controlled Trial and other specific clinical-trial types	Original Research	112
Government Publication	Roadmap/Report	105
Clinical Study, Observational Study, Comparative Study, Evaluation Study, Validation Study, Multicenter Study, Twin Study	Original Research	86
Review	Narrative Review	72

Table 2: Evaluation of PMC JATS article-type field

JATS article type	Candidate	Score
research-article, research	Original Research	62
review-article, review	Narrative Review	62
case-report	Case Report/Series	96
protocol, study-protocol	Study Protocol	96
editorial, article-commentary, commentary	Editorial/Commentary	96
letter, reply	Letter/Correspondence	96
conference-abstract	Conference Abstract	96
meeting-report	Roadmap/Report	96
correction, retraction	Correction/Retraction Notice	96
methods-article, technical-note	Methods/Technical Article	96

Table 3: Topic definitions, controlled metadata and PubMed title-query terms

Topic	Lipedema qualifiers	Topic MeSH descriptors or publication types	titleQueryTerms
Genetics & Genomics	genetics (Q000235)	Genetic Variation (D014644); Genetic Association Studies (D056726); Genome-Wide Association Study (D055106); Exome Sequencing (D000073359); Gene Expression Profiling (D020869)	genetic*[Title];genomic*[Title];genotype*[Title];variant*[Title];mutation*[Title];polymorphism*[Title];heritab*[Title];inherit*[Title];familial[Title];pedigree[Title];segregation[Title];allele*[Title];locus[Title];loci[Title];GWAS[Title];"whole exome sequencing"[Title];"exome sequencing"[Title];"whole genome sequencing"[Title];transcriptom*[Title];"RNA-seq"[Title];"RNA sequencing"[Title];"single-cell"[Title];epigenom*[Title]
Pathophysiology	physiopathology (Q000503); pathology (Q000473); immunology (Q000276); metabolism (Q000378)	Inflammation (D007249); Fibrosis (D005355); Extracellular Matrix (D005109)	pathophysiolog*[Title];mechanism*[Title];inflamm*[Title];fibros*[Title];"extracellular matrix"[Title];cytokine*[Title];macrophage*[Title];hypoxia[Title];"oxidative stress"[Title]
Hormones & Estrogen	None	Estrogens (D004967); Estradiol (D004958); Progesterone (D011374); Menopause (D008593)	hormon*[Title];estrogen*[Title];oestrogen*[Title];estradiol[Title];progesterone[Title];menopause[Title];perimenopause[Title];puberty[Title];menarche[Title];menstrual[Title];pregnancy[Title];endocrine[Title]
Diagnosis & Clinical	diagnosis (Q000175); diagnostic imaging (Q000178); classification (Q000145)	Diagnosis, Differential (D003937); Ultrasonography (D014463); Magnetic Resonance Imaging (D008279)	"diagnostic criteria"[Title];"differential diagnosis"[Title];diagnos*[Title];staging[Title];classification[Title];phenotyp*[Title];"diagnostic imaging"[Title];ultrasonograph*[Title];ultrasound[Title];MRI[Title];"diagnostic biomarker"[Title]
Treatment & Surgery	therapy (Q000628); surgery (Q000601); rehabilitation (Q000534)	Lipectomy (D008029); Compression Therapy (D019245); Exercise Therapy (D005081); Diet Therapy (D004032)	liposuction[Title];lipectomy[Title];surg*[Title];"compression therapy"[Title];"manual lymphatic drainage"[Title];physiotherap*[Title];"exercise intervention"[Title];"dietary intervention"[Title];pharmacotherap*[Title];"treatment outcome"[Title];therap*[Title]
Lymphatics & Vasculature	None	Lymphatic Vessels (D008208); Lymphatic System (D008207); Microcirculation (D008895); Lymphedema (D008209)	"lymphatic vessel"[Title];lymphangiogenesis[Title];"lymphatic dysfunction"[Title];"vascular dysfunction"[Title];endotheli*[Title];microvascular[Title];capillar*[Title];microcirculation[Title];

Topic	Lipedema qualifiers	Topic MeSH descriptors or publication types	titleQueryTerms
			lymphedema[Title]; lymphoedema[Title]
Adipose Biology	None	Adipose Tissue (D000273); Adipocytes (D017667); Lipid Metabolism (D008047); Stromal Vascular Fraction (D000090002)	"adipose tissue"[Title]; adipocyte*[Title]; adipogenesis[Title]; lipogenesis[Title]; lipolysis[Title]; "subcutaneous fat"[Title]; "stromal vascular fraction"[Title]; "adipose progenitor"[Title]
Epidemiology & Registry	epidemiology (Q000453)	Prevalence (D015995); Incidence (D015994); Registries (D012042)	prevalence[Title]; incidence[Title]; epidemiolog*[Title]; registry[Title]; registries[Title]; "population survey"[Title]; "demographic distribution"[Title]
Guidelines & Consensus	None	Publication types: Guideline (D016431); Practice Guideline (D017065)	guideline*[Title]; consensus[Title]; Delphi[Title]; "position statement"[Title]; "clinical recommendation"[Title]; "best practice recommendation"[Title]
Comorbidities & QoL	complications (Q000150); psychology (Q000523)	Quality of Life (D011788); Comorbidity (D015897)	"quality of life"[Title]; "patient-reported outcome"[Title]; comorbid*[Title]; "pain burden"[Title]; "psychological health"[Title]; depression[Title]; anxiety[Title]; "mobility limitation"[Title]; disability[Title]; fatigue[Title]; "symptom burden"[Title]

Table 4: Evidence-source logic in primary research-topic classifier. (*n*) is the number of distinct configured expressions that match. If the same expression occurs several times, it still contributes only one match.

Evidence source	Input examined	Eligibility and matching behaviour	Proposed score	Specificity	Stored source label
Disease evidence	MeSH descriptors, title and abstract	True when a MeSH descriptor label equals Lipedema, or when the case-insensitive whole word lipedema or lipoedema occurs in the title or abstract	No candidate score by itself	—	Not stored separately
Lipedema MeSH qualifier	Qualifiers attached to a MeSH descriptor labelled Lipedema	The qualifier label must equal a qualifier configured for the topic. Labels are compared case-insensitively; identifiers are not used for matching	120 if the qualifier has MajorTopicYN="Y"; otherwise 112	6	mesh-qualifier
Joint major MeSH evidence	Lipedema descriptor and a configured topic descriptor	At least one Lipedema descriptor and the matching topic descriptor must both have MajorTopicYN="Y"	104	5	major-mesh
Major topic descriptor with disease evidence	Configured topic descriptor	The topic descriptor is marked major, but the joint-major condition is not met; other disease evidence must be present	78	4	mesh-descriptor
Non-major topic descriptor with disease evidence	Configured topic descriptor	The topic descriptor is not marked major; disease evidence must be present	60	3	mesh-descriptor
Configured publication type	PubMed publication-type labels	The label must equal a publication type configured for the topic, and disease evidence must be present. Only Guidelines & Consensus has publication types configured	100	5	publication-type
Unique topic-query membership	topicQueryMembership for the PMID	The PMID belongs to the current topic query and belongs to exactly one configured topic query	90	4	exact-query-membership
Overlapping topic-query membership	topicQueryMembership for the PMID	The PMID belongs to the current topic query and belongs to more than one configured topic query	54 for every matched topic	2	overlapping-query-membership
Direct title patterns	Publication title	Every configured title regular expression is tested once. Repetition of the same expression does not increase the match count	$\min(88, 68 + 6n)$	4	title
Author-keyword patterns	PubMed author keywords joined with spaces	Tested against the combined title and ordinary abstract patterns; each distinct matching expression counts once	$\min(76, 58 + 4n)$	3	author-keywords
Indexed-substance patterns	PubMed chemicals and supplementary concepts joined with spaces	Tested against the combined title and ordinary abstract patterns; each distinct matching expression counts once	$\min(68, 50 + 4n)$	3	indexed-substances
Structured abstract section (Tables 4A-4B)	Combined text from the corresponding abstract-section labels	Requires at least one strong pattern or, when no strong pattern is present, at least two ordinary patterns	Base + 8 for any strong match; otherwise $\min(\text{base} + 12, \text{base} + 3n)$	Depends on section	Section-specific label
Stored full-text section (Tables 4A-4B)	Stored full-text Results, Conclusions or Methods fields	Uses the same strong/ordinary rule as structured sections; operates only when these fields are already populated	Base + pattern adjustment	Depends on section	Section-specific label

Evidence source	Input examined	Eligibility and matching behaviour	Proposed score	Specificity	Stored source label
Complete-abstract-fallback (Table 4C)	Full abstract after future-recommendation filtering	Evaluated separately for a topic only when that topic has received no evidence from any earlier source	48 for one or more strong patterns; otherwise $\min(44, 30 + 5n)$ when at least two ordinary patterns match	1	abstract-fallback

Table 4A: Structured-section inputs and base scores

Stored evidence label	Section-label substrings	Future-recommendation filtering	Base score	Specificity
abstract-objective-conclusion	OBJECTIVE, AIM, PURPOSE, CONCLUSION, INTERPRETATION	Yes	62	3
abstract-results	RESULT	Yes	60	3
abstract-methods	METHOD, DESIGN, MATERIAL	No	48	2
abstract-background	BACKGROUND, INTRODUCTION	Yes	38	1
full-text-results	Stored Results or labels containing RESULT	Yes	58	3
full-text-conclusions	Stored Conclusions or labels containing CONCLUSION or DISCUSSION	Yes	54	2
full-text-methods	Stored Methods or labels containing METHOD, DESIGN or MATERIAL	No	44	2

Table 4B: Resulting section scores

Source	One or more strong patterns	Two ordinary patterns	Three ordinary patterns	Four or more ordinary patterns
Objective/conclusion	70	68	71	74
Results	68	66	69	72
Methods	56	54	57	60
Background	46	44	47	50
Stored full-text Results	66	64	67	70
Stored full-text Conclusions	62	60	63	66
Stored full-text Methods	52	50	53	56

Table 4C: Future-recommendation filtering. Before selected sections and the fallback abstract are searched, LipoScope removes sentences containing the programmed expressions. The rule is intended to reduce classification based only on subjects proposed for future study. Filtering is applied to Objective/Aim/Purpose/Conclusion/Interpretation, Results, Background/Introduction, stored full-text Results, stored full-text Conclusions and the fallback abstract.

Expression group	Exact forms matched
Future-work phrase	future, further, additional, prospective or subsequent, followed by research, study, studies, work or investigation
Needed phrase	is needed, are needed, will be needed
Should-be phrase	should be investigated, should be evaluated, should be studied, should be validated, should be examined

Table 5: Construction of topic-query arms. The qualifier arm is omitted when a topic has no configured Lipedema qualifier. All ten topics have a title arm and either a descriptor-based or publication-type-based Major Topic arm.

Arm	Construction
Qualifier arm	Each configured qualifier becomes "Lipedema/<qualifier>"[Mesh]; multiple qualifiers are joined using OR
Major Topic arm	("Lipedema"[Majr] AND (<descriptor>[Majr] OR <publication type>[Publication Type] ...))
Title arm	((lipedema[Title] OR lipoedema[Title]) AND (<titleQueryTerm> OR ...))
Complete topic query	All available arms are enclosed in parentheses and joined using OR

Table 6: Direct title expressions

Topic	Direct title expressions accepted
Genetics & Genomics	Words beginning genetic, genomic, genotype, variant, mutation, polymorphism, heritab, inherit, allele, transcriptom or epigenom; exact whole words familial, pedigree, segregation, GWAS; locus or loci; whole-exome, whole exome or exome sequencing; whole-genome or whole genome sequencing; RNA-seq, RNA seq, RNA-sequencing or RNA sequencing; single-cell or single cell
Pathophysiology	Words beginning pathophysiolog, mechanism, inflamm, fibros, cytokine or macrophage; exact phrases extracellular matrix, oxidative stress; exact word hypoxia
Hormones & Estrogen	Words beginning hormon, estrogen or oestrogen; exact words estradiol, progesterone, menopause, perimenopause, puberty, menarche, menstrual, pregnancy; exact phrase endocrine signalling
Diagnosis & Clinical	Exact phrases diagnostic criteria, differential diagnosis, diagnostic imaging; words beginning diagnos, phenotyp, ultrasonograph or diagnostic biomarker; exact words staging, classification, ultrasound, MRI
Treatment & Surgery	Exact words liposuction, lipectomy; words beginning surg, physiotherap, pharmacotherap or treatment outcome; exact phrases compression therapy, manual lymphatic drainage, exercise intervention, dietary intervention, therapeutic intervention
Lymphatics & Vasculature	lymphatic vessel followed by zero or more word characters; exact phrases lymphatic dysfunction, vascular dysfunction; words beginning endotheli or capillar; exact words lymphangiogenesis, microvascular, microcirculation, lymphedema, lymphoedema
Adipose Biology	Exact phrases adipose tissue, subcutaneous fat, stromal vascular fraction; words beginning adipocyte or adipose progenitor; exact words adipogenesis, lipogenesis, lipolysis
Epidemiology & Registry	Exact words prevalence, incidence; words beginning epidemiolog; exact words registry or registries; exact phrases population survey, demographic distribution
Guidelines & Consensus	Words beginning guideline, clinical recommendation or best practice recommendation; exact words consensus, Delphi; exact phrase position statement
Comorbidities & QoL	Exact phrase quality of life; patient-reported or patient reported outcome followed by zero or more word characters; words beginning comorbid or disabilit; exact phrases pain burden, psychological health, mobility limitation, symptom burden; exact words depression, anxiety, fatigue

Table 7: Ordinary and strong abstract patterns. LipoScope uses ordinary and strong expressions for structured abstract sections, stored full-text sections and complete-abstract fallback. A strong pattern is also present in the topic's ordinary pattern list but the classifier gives it special treatment. One strong match is sufficient for a section to provide evidence. When no strong pattern is present, at least two distinct ordinary patterns are required.

Topic	Ordinary abstract patterns (abstractPatterns)	Strong abstract patterns (strongAbstractPatterns)
Genetics & Genomics	genome-wide association study/studies; whole-exome or whole exome sequencing; whole-genome or whole genome sequencing; rare variant/variants; rare coding variant/variants; pathogenic variant/variants; genetic association; segregation analysis; RNA-sequencing or RNA sequencing; transcriptomic analysis	genome-wide association study/studies; whole-exome sequencing; whole-genome sequencing; segregation analysis
Pathophysiology	chronic inflammation; extracellular-matrix remodeling/remodelling; immune-cell infiltration; macrophage polarization; tissue fibrosis; hypoxic signalling; oxidative stress	extracellular-matrix remodeling/remodelling; macrophage polarization
Hormones & Estrogen	estrogen receptor signalling; oestrogen receptor signalling; progesterone metabolism; reproductive hormone/hormones; menopausal transition; hormonal transition	estrogen receptor signalling; oestrogen receptor signalling; progesterone metabolism
Diagnosis & Clinical	diagnostic accuracy; differential diagnosis; diagnostic criteria; clinical classification; imaging biomarker; magnetic resonance imaging	diagnostic accuracy; diagnostic criteria
Treatment & Surgery	liposuction outcome/outcomes; surgical outcome/outcomes; compression garment/garments; manual lymphatic drainage; conservative treatment; therapeutic intervention	liposuction outcome/outcomes; surgical outcome/outcomes; manual lymphatic drainage
Lymphatics & Vasculature	lymphatic-vessel dysfunction; lymphatic-drainage impairment; endothelial dysfunction; microvascular permeability; capillary fragility; altered microcirculation	lymphatic-vessel dysfunction; endothelial dysfunction; microvascular permeability
Adipose Biology	adipocyte differentiation; adipose-tissue remodeling/remodelling; adipocyte hypertrophy; stromal vascular fraction; adipose progenitor cell/cells; lipid-storage dysfunction	adipose-tissue remodeling/remodelling; stromal vascular fraction; adipose progenitor cell/cells
Epidemiology & Registry	population prevalence; population-based prevalence; incidence rate; national registry; population survey; demographic distribution	population-based prevalence; national registry
Guidelines & Consensus	consensus statement; Delphi process; clinical practice guideline; evidence-based recommendation/recommendations	consensus statement; Delphi process; clinical practice guideline
Comorbidities & QoL	health-related quality of life; patient-reported outcome/outcomes; psychological distress; pain severity; mobility limitation; symptom burden	health-related quality of life; patient-reported outcome/outcomes; psychological distress

Table 8: Combining evidence for one topic. For each topic, LipoScope begins with a score and specificity of zero. Every successful evidence rule records an evidence entry. The classifier retains the largest individual source score and, separately, the largest specificity tier. It does not sum all original source scores. The final topic score is: $\min(125, \max(\text{individual source scores}) + \min(12, 3(\text{number of evidence entries} - 1)))$. The practical effect of the corroboration bonus is to give limited additional weight when a topic is supported in more than one place. For example, agreement between a MeSH descriptor, a title expression and an abstract expression strengthens the topic candidate. The bonus is deliberately capped in the implementation, so the presence of multiple evidence entries cannot increase the score without limit.

Operation	Code behaviour
Initial score	Starts at zero separately for every topic
Adding evidence	score = max(current score, proposed source score)
Specificity	specificity = max(current specificity, proposed source specificity)
Evidence provenance	Every successful evidence-source call appends an evidence entry
Corroboration	Three points per evidence entry beyond the first
Corroboration cap	12 points
Overall topic-score cap	125
Candidate creation	A topic becomes a candidate whenever it has at least one evidence entry
Independent maxima	Maximum score and maximum specificity can originate from different sources
Original source scores	Not added together

Table 9: Comparing candidate topics. Every topic with at least one evidence entry becomes a candidate. Candidates are compared using the order and acceptance rules. Because specificity is evaluated before score, a candidate with greater specificity can be placed first even when another candidate has a larger numerical score. Priority is considered only when specificity and score are equal.

Stage	Rule
First ordering criterion	Descending specificity
Second ordering criterion	Descending final topic score
Third ordering criterion	Descending fixed topic-priority value (Table 9A)
Minimum score	Leading candidate must have score ≥ 58
Acceptance route 1	No second candidate exists
Acceptance route 2	The leader has greater specificity than the second candidate
Acceptance route 3	The leader exceeds the second candidate by at least 8 score points
Acceptance route 4	The leader contains exact-query-membership evidence
Acceptance route 5	The leader has specificity ≥ 5 and exceeds the second candidate by at least 3 points
Abstention	If the minimum score or all separation routes fail, return Unsupported records
Conflict reason	multiple-equivalent-topic-evidence

Table 9A: Fixed deterministic topic-priority order. Priority is not added to an evidence score and does not indicate scientific importance, research priority or classification accuracy.

Topic	Priority
Guidelines & Consensus	1000
Genetics & Genomics	900
Treatment & Surgery	800
Diagnosis & Clinical	700
Epidemiology & Registry	600
Comorbidities & QoL	550
Lymphatics & Vasculature	500
Adipose Biology	450
Hormones & Estrogen	400
Pathophysiology	300

Table 10: Final classification output. When the leading candidate satisfies the programmed rules, it becomes the record's primary topic. Each record contributes to no more than one primary-topic count. The classifier output fields are summarized in this table. The confidence values are code-defined score bands. When a primary topic is accepted, non-primary candidates scoring at least 45 are retained as secondary topics. They are not included as additional primary-topic counts. When classification abstains because candidates cannot be separated, all candidate names are stored in secondaryTopics, without applying the score-45 filter; the primary result remains Unsupported records.

Output field	Behaviour
topic and primaryTopic	Accepted topic name or Unsupported records
score	Final leader score; zero when no candidate exists
confidence	High at ≥90; moderate at 68–89; low below 68
evidence	Evidence supporting the accepted topic; unsupported or abstention explanation when no topic is accepted
evidenceByTopic	Score, specificity and evidence entries for every generated topic candidate
matchedTopics	Other candidates with their scores and specificity tiers
secondaryTopics	For an accepted classification, non-primary candidates with scores ≥45
unsupported	false for an accepted topic; true otherwise
unsupportedReason	Empty for accepted records; coded reason for unsupported records
classifierVersion	mesh-hybrid-multilabel-query-membership-2026.4

Table 11: Unsupported-record reason selection. If no topic satisfies the decision rules, the record is returned as Unsupported records. This table shows the exact reason-selection order.

Order	Condition	Stored reason
Conflict path	Candidates exist but fail the separation rules	multiple-equivalent-topic-evidence
1	No candidate; citation status is Publisher, In-Process or In-Data-Review	awaiting-medline-indexing
2	No candidate; both abstract and author keywords are absent	abstract-unavailable
3	No candidate; citation status is PubMed-not-MEDLINE	pubmed-not-medline
4	No candidate; MeSH, publication types and citation status are all absent	metadata-retrieval-incomplete
5	Remaining no-candidate cases	no-category-specific-evidence