

# Barriers and Enablers to Digital Health Literacy: A Thematic Analysis of Patient Interviews

Journal of Health Informatics Research · Vol. 12, No. 3, 2024 · DOI: 10.1007/jhir.2024.0312

Twenty semi-structured interviews were conducted with adult patients (aged 28–74) attending a primary care clinic in a mid-sized urban centre. Interviews were audio-recorded, transcribed verbatim, and analysed using reflexive thematic analysis (Braun & Clarke, 2006). Four overarching themes were identified, each containing two to four sub-themes. Representative quotations for each sub-theme are presented in Table 1 below. Participant identifiers (P-01 to P-20) preserve anonymity.

**Table 1. Thematic analysis: themes, sub-themes, and representative quotations**

Category rows (shaded) mark theme boundaries. Each data row corresponds to one participant quotation coded to a sub-theme.

Participant	Verbatim Quote (excerpt)	Sub-theme
Theme 1: Information Access Barriers		
P-01	"I tried to look up my test results online but the patient portal kept logging me out. I gave up after the third time."	Digital access
P-04	"My English is not perfect so I use Google Translate, but the medical words come out wrong. I am confused about what my doctor wrote."	Language & literacy
P-07	"I live in a rural area. The internet connection is very slow. Downloading the health app took two days."	Connectivity
Theme 2: Trust and Credibility		
P-03	"I found five different answers on five different websites. How am I supposed to know which one is correct?"	Information overload
P-11	"I only trust what my GP tells me. The internet is full of nonsense from people who are not doctors."	Source credibility
P-14	"After reading the side-effects list I stopped taking my medication. I wish someone had explained what was really likely to happen."	Risk perception
Theme 3: Healthcare System Navigation		
P-06	"I did not know I could ask for a second opinion. Nobody told me that was an option. I just assumed the first doctor was right."	Patient rights
P-09	"The referral process is completely opaque. I waited four months and nobody told me what was happening or why."	System transparency
P-16	"Every department seems to have a different form. I filled in my details six times for the same admission."	Administrative burden
Theme 4: Emotional and Social Dimensions		
P-02	"I felt embarrassed to admit I did not understand the diagnosis. I just nodded and left. I cried in the car."	Emotional response

Participant	Verbatim Quote (excerpt)	Sub-theme
P-13	"My daughter helps me with all the online stuff. Without her I would be completely lost."	Social support
P-18	"When I got the diagnosis I went down a rabbit hole on YouTube for two weeks. It made my anxiety much worse."	Information-seeking behaviour

## PIPELINE FAILURE ANALYSIS

**Expected behaviour:** Table 1 preserved with 4 category rows and 12 data rows intact. Category rows (e.g. 'Theme 1: Information Access Barriers') appear as standalone rows that visually separate thematic groups. Final DataFrame: 16 rows × 3 cols (header + 4 categories + 12 data rows).

**What the pipeline actually produces:** `stitch_split_cells()` folds every category row into the Verbatim Quote cell of the data row that precedes it. Category rows have exactly 1 non-empty cell (col 0 = theme label, cols 1–2 = empty) — the same pattern as a genuine split-cell continuation. The function cannot distinguish them. Result: 'Theme 2: Trust and Credibility' gets appended to P-07's quote; 'Theme 3: Healthcare System Navigation' gets appended to P-14's quote; etc. The grouping structure of the table is silently destroyed. Final DataFrame: 12 rows × 3 cols (4 rows fewer, labels embedded in quotes).

### Code path responsible:

`merger.py` `stitch_split_cells()` lines 369–427. The loop at line 394 checks only: `len(nonempty_idx) == 1`. There is no check for cell position (first column vs last), row height, content length, or whether the non-empty value is long enough to be a label rather than overflow text. Fix candidate: skip folding if the non-empty cell is in column 0 AND the cell text is longer than a configurable threshold (e.g. 20 chars), OR if the cell text does not share any tokens with the previous row's cells.