

Article

Not peer-reviewed version

Documenting Concordance: A Scoping Review of NLP and Text Mining in Oncology EHR Narratives

[Leon Wreyford](#)*, [Raj Gururajan](#), [Xujuan Zhou](#), [Niall Higgins](#)

Posted Date: 11 February 2026

doi: 10.20944/preprints202602.0751.v1

Keywords: adverse effects; documentation quality; electronic health records; nursing informatics; oncologist notes; person-centred care; stigma in notes; text mining; treatment concordance



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Documenting Concordance: A Scoping Review of NLP and Text Mining in Oncology EHR Narratives

Leon Wreyford ^{1,*}, Raj Gururajan ¹, Xujuan Zhou ¹ and Niall Higgins ^{1,2}

¹ School of Business, University of Southern Queensland, Springfield Central, QLD 4300, Australia

² The Park - Mental Health and Specialised Services, West Moreton Health, Wacol, QLD, Australia

* Correspondence: u1098304@umail.usq.edu.au

† These authors contributed equally to this work: L.W.

Abstract

Background: Documentation in oncologist Electronic Health Records (EHRs) plays a critical role in communication, shared decision-making, and the detection of adverse effects all of which influence treatment concordance and adherence. However, narrative content is often incomplete, delayed, or written in formal styles that obscure patient priorities. **Methods:** Following PRISMA-ScR guidelines, we conducted a scoping review of studies published between 2013 and early 2024 that used natural language processing (NLP) or text mining on oncologist notes, or qualitatively examined EHR use in oncology. Data were charted by topic (e.g., adverse effects, note style, stigma, workflow burden) and synthesized using discourse analysis. **Results:** Twenty-three studies met inclusion criteria. Four clinician-side themes emerged: (1) compliance-oriented EHR design; (2) incomplete or delayed documentation of adverse effects, pain, and social determinants of health (SDOH); (3) formal or stigmatizing language; and (4) time and cognitive burden limiting person-centred narratives. These factors hinder concordance documentation and communication transparency. **Conclusions:** Improvements in EHR design, including person-centred prompts, plain-language templates, and audit-log-informed workflow changes, may enhance adherence by supporting better documentation and communication. Nursing and service leaders can implement these changes to promote trust, engagement, and continuity of care. **Registration:** Not registered.

Keywords: adverse effects; documentation quality; electronic health records; nursing informatics; oncologist notes; person-centred care; stigma in notes; text mining; treatment concordance

1. Introduction

Electronic Health Records (EHRs) are central to oncology care, yet much of their clinically relevant content is captured in unstructured text progress notes, pathology/radiology reports, and other narrative fields [1,2]. Narrative documentation is the locus where adverse effects, symptom trajectories, psychosocial issues, patients' goals and values, and shared decision making should be recorded. However, multiple studies suggest that such information is inconsistently captured, delayed, or written in formal styles that obscure patient priorities [3,4]. Access and data sharing constraints, privacy concerns, and limited availability of shared de-identified corpora have also slowed oncology specific Natural Language Processing (NLP) [5,6]. Recent reviews note that extracting oncology concepts from clinical notes remains an active methodological frontier [7,8].

Documentation gaps have implications for treatment concordance and adherence. Adverse effects and pain may be incompletely recorded, and social determinants of health (SDOH) under documented, hindering timely recognition of barriers to ongoing therapy [1,9–11]. Language choices in notes can transmit bias, with stigmatizing terms linked to alienation and trust erosion [12]. At the same time, clinicians face time pressure and cognitive burden: template driven and compliance oriented EHRs demand substantial navigation and coding effort, which can crowd out person-centred narratives,

encourage copy paste practices, and limit explicit recording of concordance conversations [13–18]. The result is a documentation ecosystem optimized for billing and standardized data capture, but not always for the transparency, inclusivity, and specificity that shared decision making requires [19,20].

NLP and text mining make it possible to analyze clinician notes at scale, revealing patterns in documentation density, topic emphasis, language tone, and reporting delays. Studies have used such methods to classify serious illness communication and to compare recorded notes with the substance of clinical encounters, highlighting discrepancies and missed opportunities to codify patient goals and agreements [21,22]. Similar research on patient access to notes underscores the communicative function of documentation itself. Clinical notes are often in plain language and person-centred phrasing can strengthen understanding, trust, and engagement [4,23,24].

Aim and research questions. This scoping review synthesizes evidence from NLP and text mining studies of oncologist EHR narratives. Qualitative reports on EHR use in oncology were also included because they can also identify documentation patterns that may impede or enable treatment concordance and adherence. Using discourse analysis, we ask:

1. Which documentation patterns (e.g., adverse effect reporting, pain, SDOH capture, note style, lexicon, timing, reuse) hinder the recording and retrieval of information that is relevant to concordance [9,10,12]?
2. How can note language and workflow design (e.g., prompts, concordance fields, audit-log informed process changes) be optimized to support person-centred care and improve adherence outcomes [13,14,20]?

2. Method

2.1. Design and Reporting

We conducted a scoping review compliant with PRISMA-ScR [25] to map NLP and text mining evidence from oncologist EHR narratives and qualitative EHR studies in oncology. The goal was to identify documentation patterns linked to concordance and adherence, and to derive nursing service design implications.

2.2. Eligibility Criteria

Inclusion. Studies were eligible if they: (i) mined oncologist or clinician notes or oncology clinical documentation using NLP or text mining; or (ii) investigated oncologist EHR use qualitatively with implications for documentation and communication; (iii) reported on constructs relevant to concordance and adherence; (iv) were peer reviewed and published in English between 2013 and early 2024 [4,7,8,21].

Exclusion. We excluded non oncology note mining, non English publications, economic analyses, and grey literature except World Health Organisation (WHO) and Australian Commission on Safety and Quality in Health Care (ACSQHC) policy statements used for definitional clarity.

2.3. Information Sources and Search Strategy

We searched PubMed, CINAHL, and Scopus (2013–early 2024) using combinations of oncology and EHR terms and documentation constructs (e.g., “oncology” AND “electronic health records” AND “clinical notes”; “text mining” AND “adverse effects” AND “oncology”; “stigma” AND “clinical documentation”). The strategy was iteratively refined to include design and workflow terms (e.g., audit logs, template burden, copy–paste) and language and style concepts (plain language, person-centred phrasing, stigmatizing lexicon) [12–14,20]. We noted barriers to corpus access and de-identification highlighted in informatics studies [5,6] and the infancy of oncology specific NLP [7].

2.4. Study Selection

Two reviewers independently screened titles and abstracts for EHR narrative mining or oncology EHR relevance, then examined full texts against inclusion and exclusion criteria. Disagreements were resolved by consensus. The process will be documented in a PRISMA-ScR flow diagram (Figure 1) [25].

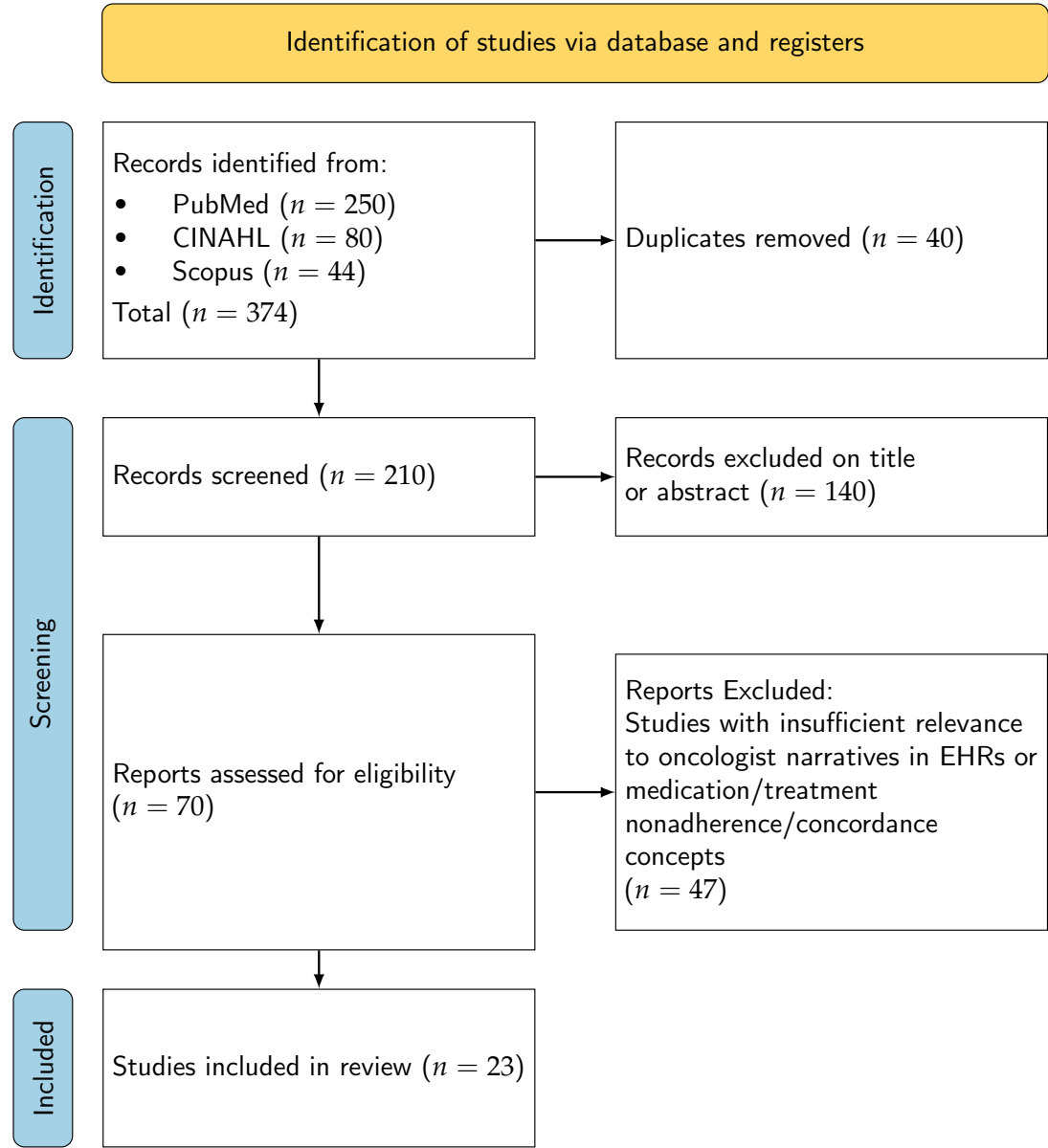


Figure 1. PRISMA Flow Diagram [25].

2.5. Data Charting

We charted author, year, and country, cancer domain, EHR system context, corpus properties (size, note types), NLP and text mining methods, targeted documentation topics (e.g., adverse effects, pain, SDOH, concordance cues, note style), and key findings. We also captured feasibility indicators (time and cognitive burden, reuse or copy–paste, template constraints), and language considerations (plain language phrasing, stigma markers, person-centred content) [12,15,20].

2.6. Critical Appraisal

We used the Critical Appraisal Toolkit (CAT) to describe the methodological rigor of heterogeneous sources (NLP pipelines; qualitative studies; mixed-methods) with ratings (low/medium/high) and strength of evidence descriptors [26]. Appraisal results informed interpretation; consistent with scoping review methodology, studies were not excluded solely on appraisal scores [27,28].

2.7. Synthesis Approach

Discourse analysis guided a narrative synthesis of clinician side themes: (i) EHR barriers or compliance-oriented design; (ii) incomplete or late documentation (adverse effects, pain, SDOH); (iii) formal note style and stigmatizing language; and (iv) time and cognitive burden limiting Patient

Centred Care (PCC) narratives [12–14,20]. We used sentinel exemplars to illustrate misalignment between recorded notes and serious illness communication or patient goals [21,22], and integrated evidence on patient access portals to emphasize documentation’s communicative function [4,23,24]. A thematic barrier map (Figure 2) summarizes the relationships between documentation features and concordance and adherence outcomes.

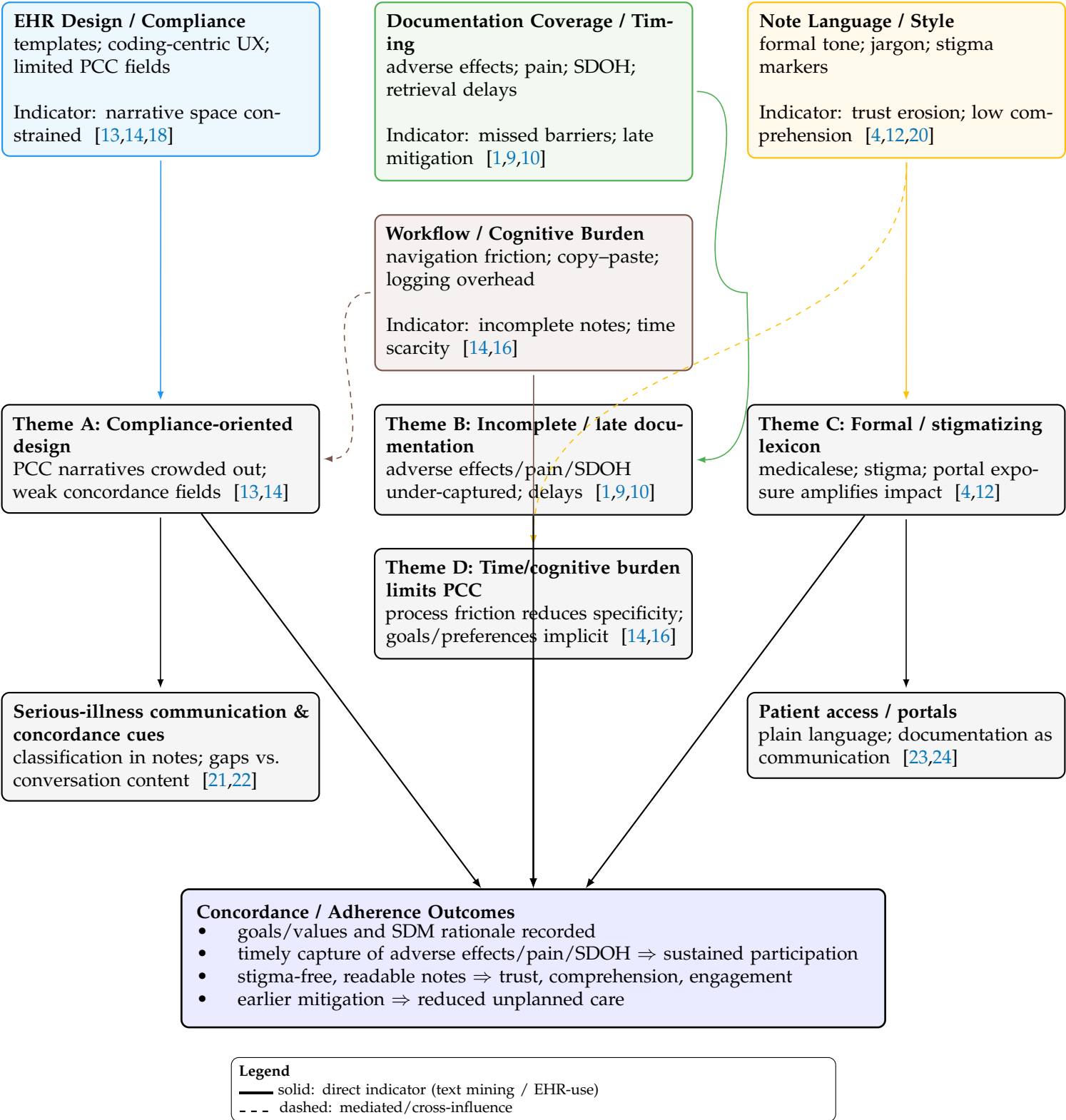


Figure 2. Conceptual links between EHR determinants and clinician-side themes (barriers, coverage/timing, language/style, workflow burden) and their effects on concordance/adherence outcomes [1,4,9,10,12–14,16,20]; bridging constructs: [21–24]).

3. Results

3.1. Overview

A total of twenty–three studies met the inclusion criteria for this scoping review and were synthesized to characterize documentation patterns in oncology EHR narratives and their implications for treatment concordance and adherence. Reasons for exclusion at full text stage comprised non oncology note mining, non English publications, economic analyses, and grey literature beyond WHO and ACSQHC definitional sources. The overall selection process is summarized in the PRISMA–ScR flow diagram (Figure 1) [25].

3.2. Study Selection Narrative (PRISMA–ScR)

Across database searches (PubMed, CINAHL, Scopus) and iterative query refinement, records were screened on title and abstract for EHR narrative mining or oncology EHR relevance, followed by full text assessment against predefined inclusion/exclusion criteria. Disagreements between reviewers were resolved by consensus. While exact counts at each PRISMA node are depicted in Figure 1, the final corpus comprises 23 included studies spanning NLP/text mining of oncologist notes and qualitative examinations of oncology EHR use [4,7–10,12–14,20–23].

3.3. Characteristics of Included Studies

Table 1 summarizes the most salient features of the included sources: cancer domain, corpus type and size, EHR context, analytic method (e.g., rule based extraction, machine learning, linguistic analysis), and concordance and adherence relevant indicators derived from narrative documentation. Across studies, corpora and pipelines were heterogeneous ranging from progress notes and discharge letters to audit logs and portal view data yet converged on documentation constructs central to concordance (adverse effects, pain, SDOH, note style, timing, reuse) [1,4,9,10,12–15,20].

Table 1. Summary of oncologist studies (Part 1 of 3)

Author & Title	Method	Data Sources	Sample	Results	NLP Technique	Nonadherence/Inference
Masukawa et al., 2022 [1] . Machine Learning Models to Detect Social Distress and severe physical/psychological symptoms in terminally ill patients with cancer from unstructured EHR text data	Quantitative study	Retrospective cancer patient records	808 cancer patients. CAT: Strength=Moderate; Quality = Medium	ML detected social distress and symptoms (pain, dyspnea, nausea, insomnia, anxiety).	Supervised machine learning	Text data can identify social distress; areas like anxiety receive scant attention compared to physical pain (~40% of narratives), affecting adherence risks.
Alpert et al., 2019 [4] . Patient access to clinical notes in oncology: A mixed method analysis of oncologists’ and linguistic attitudes towards notes	Mixed methods	National Cancer Institute, Virginia	13 interviews; 500 clinical notes; 22 oncologists. CAT: Strength=Moderate; Quality = Medium	Oncologists acknowledged that changing note content could improve patient communication but may hinder interdisciplinary communication.	Linguistic Inquiry and Word Count (LIWC); random effects modelling	Tension between patient-centred language and clinical interaction; challenges in non-clinical communication may affect adherence.
Che-Chen Kuo et al., 2022 [9] . Using data mining technology to predict medication-taking behaviour in women with breast cancer: A retrospective study	Quantitative (structured EHR data)	Breast cancer patient records	385 records. CAT: Strength = Weak; Quality = Low	Highest polarity of reviews related to patient–doctor experience and pain.	Multiple logistic regression, decision tree, artificial neural network	Structured medical records may not capture adverse effects; missing factors can influence adherence.

Table 1. Cont.

Author & Title	Method	Data Sources	Sample	Results	NLP Technique	Nonadherence/Inference
Tamang et al., 2015 [10] . Detecting Unplanned Care from Clinician Notes in Electronic Health Records	Quantitative study	Cancer patient EHRs	308,096 free-text machine-readable notes. CAT: Strength=Moderate; Quality = Medium	Including free-text notes increased identification of ED visits; textual analysis identified most reported symptoms.	Clinical text mining	Combining structured & unstructured data improved detection; pain was the most common reason for ED visits — cancer pain management suboptimal.
Himmelstein et al., 2023 [12] . Examination of Stigmatising Language in the Electronic Health Record	Quantitative	Longitudinal retrospective	National Danish Patient Registry; 4,418 pancreatic cancer patients. CAT: Strength=Moderate; Quality = Medium	Text-based approach identified 132 unique symptoms in clinical notes.	Text mining comparisons to registry data	Stigmatizing language appears in admission notes and varies by condition; stigmatization may alienate patients and lead to treatment discordance.
Patel et al., 2023 [13] . Clinician Perspectives on Electronic Health Records, Communication, and Patient Safety Across Diverse Medical Oncology Practices	Mixed-methods	Michigan Oncology Quality Collaborative (MOQC)	29 oncology practices; 297 clinicians. CAT: Strength=Moderate; Quality = Medium	More seamless EHR integration into routine care is needed.	Sociotechnical framework; surveys	EHRs disrupt communication and increase workloads; patient communication may be compromised, risking unintentional nonadherence.
Huilgol et al., 2022 [14] . Opportunities to use electronic health record audit logs to improve cancer care	Qualitative	Report (article)	Summary on interpreting metadata from oncologist audit logs. CAT: Strength=Moderate; Quality = Medium	Audit logs can show how oncologists access information and collaborate.	Non-NLP (audit log analytics)	Understanding audit logs may clarify barriers to communication from EHR dissatisfaction (time/data entry/navigation), with adherence implications.
Rahimian et al., 2019 [15] . Significant and Distinctive n-grams in Oncology Notes: A Text-Mining Method to Analyse the Effect of OpenNotes on Clinical Documentation	Quantitative (n-gram analysis)	Oncology provider notes	102,135 notes by 36 clinicians. CAT: Strength=Moderate; Quality = Medium	Significant differences before/after patient access to notes; longer explanatory notes observed.	Text mining (n-grams)	Oncologists may be less prolific note-takers; subtle empathetic changes with patient access may reflect communication skills relevant to adherence.

Table 2. Summary of oncologist studies (Part 2 of 3)

Author & Title	Method	Data Sources	Sample	Results	NLP Technique	Nonadherence/Inference
Asan et al., 2023 [17] . Oncologists’ views regarding the role of Electronic Health Records in Care Coordination	Qualitative report	Semi-structured interviews	60 oncologists. CAT: Strength=Moderate; Quality = Medium	Perceptions of oncologist EHR use during care coordination.	Traditional qualitative (non-NLP)	Oncologists report EHRs often not updated during encounters, potentially impacting communication and adherence.
Geskey et al., 2023 [19] . National Documentation and Coding Practices of Noncompliance: The Importance of Social Determinants of Health and the Stigma of African American Bias	Quantitative	Z codes (social determinants) in U.S. patient records	~9 million adult patients. CAT: Strength = Medium; Quality = Medium	Two or three comorbidities doubled the ratio of noncompliance.	Analysis of ICD-10 diagnosis codes in EHRs	Clinician bias in noncompliance coding (e.g., financial circumstances) may misclassify nonadherence from a multifactorial perspective.

Table 2. Cont.

Author & Title	Method	Data Sources	Sample	Results	NLP Technique	Nonadherence/Inference
Davoudi et al., 2022 [21] . Using Natural Language Processing to Classify Serious Illness Communication with Oncology Patients	Quantitative study	University of Pennsylvania Abramson Cancer Center	3,563 patients; 5,145 notes; 8,695 distinct responses. CAT: Strength=Moderate; Quality = Medium	Validated SIC classifier enabling quality metrics (communication quality, goal-concordance).	Machine learning (domain/sub-domain classification)	NLP can identify treatment concordance, a potential factor in nonadherence.
Simoulin et al., 2023. [36] From free-text electronic health records to structured cohorts: Onconum, an innovative methodology for real-world data mining in breast cancer	Quantitative study	Hospital patients with breast cancer at Strasbourg University Hospital	9,599 patients. CAT: Strength = Weak; Quality = Low	Successfully extracted and structured information from EHRs in breast cancer without pre-existing dictionaries or manually annotated corpora.	Hybrid NLP (machine learning + rule-based lexical methods)	Multiple sources improved extraction rates and quality; structured information can be recovered from text. May enable comparison of codified adherence against notes.
Xin et al., 2023.[37] Pancreatic cancer symptom trajectories from Danish registry data and free text in electronic health records	Quantitative	Longitudinal retrospective	National Danish Patient Registry; 4,418 pancreatic cancer patients. CAT: Strength=Moderate; Quality = Medium	Text-based approach identified 132 unique symptoms in clinical notes.	Text mining comparisons to registry data (technique not stated)	Text mining identified more symptoms than coded registries alone, potentially improving detection of issues related to adherence.
Elbers et al., 2023.[38] Sentiment analysis of medical record notes for lung cancer patients at the Department of Veterans Affairs	Quantitative study	Department of Veterans Affairs (U.S.) lung cancer patients	10,000 patients; 3,500,000+ notes. CAT: Strength=Moderate; Quality = Medium	Recalibration of a general-purpose sentiment lexicon for oncology.	Hedonometer sentiment scoring adapted to clinical oncology	Detected higher adverse effects in final days of treatment/chemotherapy, suggesting increased risk of nonadherence toward end of cycles.
Leis et al., 2022.[39] Exploring the association of cancer and depression in Electronic Health Records: Combining Encoded Diagnosis and Mining Free-Text Clinical Notes	Quantitative	EHR database in a general hospital	4,238 cancer patients. CAT: Strength=Moderate; Quality = Medium	Mining free-text identified more patients with depression than encoded data alone.	Open-source language analysis framework; text analysis	Depression often absent from structured data but detectable in notes; unintentional nonadherence may be influenced.
Mashima et al., 2022.[40] Using Natural Language Processing Techniques to Detect Adverse Events from Progress Notes Due to Chemotherapy	Quantitative	Kagawa University Hospital	200 cancer patients receiving chemotherapy. CAT: Strength=Moderate; Quality = Medium	Significant increase in adverse effect detection via progress notes analysis.	Dictionary/corpus + proprietary NLP system	Adverse effects (a factor in nonadherence) may be missed in coded data; text mining of developmental progress needed.
Jensen et al., 2017.[41] Analysis of free text in electronic health records for identification of cancer patient trajectories	Quantitative	University Hospital of North Norway	1,133,223 unstructured EHR text documents; 7,741 patients. CAT: Strength=Moderate; Quality = Medium	Free-text methods to identify trajectories may support decisions, decrease adverse events/readmissions, and improve cancer care quality.	Machine learning; text analyses using NOR-MeSH for automated capture	Time constraints for free-text entries are severe; patient access to notes may impact relationships and influence nonadherence/discordance.

Table 3. Summary of oncologist studies (Part 3 of 3)

Author & Title	Method	Data Sources	Sample	Results	NLP Technique	Nonadherence/Inference
Feldman et al., 2016.[42] Mining the Clinical Narrative: All Text are not Equal	Quantitative	Publicly available electronic medical records	45,000 patient records. CAT: Strength = Weak; Quality = Medium	Deep understanding of clinical text required; coding cannot fully describe clinician diagnosis; nuanced language is critical.	Natural Language Toolkit (NLTK) & Python	Nuanced language complicates NLP on clinical text; notes seldom record patient feedback in non-clinical terms, limiting adherence insights.
Ganesan et al., 2016.[43] Discovering Related Clinical Concepts Using Large Amounts of Clinical Notes	Quantitative study	MIMIC database	10,000 clinical notes. CAT: Strength=Moderate; Quality = Medium	Related concepts can be used for query expansion, hypothesis generation, incident investigation, sentence completion, etc.	Mining related clinical concepts with graph data structures	Related concepts in notes may be keystones for inferring treatment nonadherence.
Falotico et al., 2015.[44] Identifying Oncological Patient Information Needs to Improve e-Health Communication: a preliminary text-mining analysis	Mixed methods (quantitative text → qualitative interviews)	Semi-structured interviews; text mining	12 rare cancer patients; 20,400 words; >3,000 distinct terms. CAT: Strength=Moderate; Quality =Medium	Online communication should be supervised.	Text mining (interview transcripts)	Information needs become more treatment-specific and sophisticated over the trajectory; physician relationship quality may modulate adherence.
Hamid et al., 2020.[45] Text Parsing Based Identification of Patients with Poor Glaucoma Medication Adherence in the Electronic Health Record	Quantitative study	University of Michigan	736 glaucoma patients. CAT: Strength = Weak; Quality = Low	Identified a larger proportion of patients with poor adherence than automated EHR pull alone.	Text parsing of physician notes	Medication adherence identifiable only in progress notes, emphasizing unstructured text value.
Zhu et al., 2019.[46] Automatically identifying social isolation from clinical narratives for patients with prostate cancer	Quantitative study	Medical University of South Carolina	3,138 patients. CAT: Strength=Moderate; Quality = Medium	Highly accurate identification of social isolation when noted in clinical narratives.	NLP algorithm + lexicon for social isolation	Poor documentation of social factors in notes can lead to nonadherence; both physician and shared factors involved.
Datta et al., 2019.[47] A frame semantic overview of NLP-based information extraction for cancer-related EHR notes	Quantitative study (overview)	NLP literature	78 articles. CAT: Strength = Weak; Quality = Low	Guidance for general-purpose cancer frame resources and NLP systems to extract diverse cancer information types.	Frame construction; scoping review of NLP techniques	Treatment adherence not specifically framed despite importance; gap for future work.

3.4. Main Findings at a Glance

Synthesis of the 23 studies reveals four clinician-side themes that recur across oncology documentation and workflow research (prevalence summarized in Table 4):

- Compliance-oriented EHR design and barriers: Template burden, coding-centric UX, and immature designs constrain narrative space for person-centred care and explicit concordance fields [13,14,18,19].
- Incomplete or late documentation: Under-capture of adverse effects, pain, and SDOH, with retrieval delays that hinder timely mitigation and sustained participation [1,9–11].
- Formal or stigmatizing language: Jargon-heavy tone and stigma markers undermine trust and comprehension, especially in contexts of patient access to notes [4,12,20,23].
- Time and cognitive burden: Navigation friction, duplicative coding, and copy–paste practices reduce specificity and crowd out person-centred narratives [14,16,18].

These themes and their interrelations are depicted in the conceptual map (Figure 2), highlighting pathways by which documentation features affect concordance/adherence outcomes [1,4,9,10,12–14,16,20].

Table 4. Key clinician side themes and prevalence.

Theme	% of P2 sources
EHR barriers and compliance oriented design (templates, coding, immature UX) [13,14,19]	47%
Incomplete/late documentation (adverse effects, pain, SDOH) [1,9–11]	35%
Formal note style / stigmatizing language [4,12,20]	12%
Time/cognitive burden limiting PCC narratives [14,16,18] (frequent; overlaps with barrier theme)	

Note: Percentages reflect proportions reported in the EHR subset (Table 3). Time/cognitive burden is integral to barrier design and often co-occurs; a single percentage is therefore not reported independently.

3.5. Theme Wise Results

Theme A: Compliance Oriented Design Constrains Person Centred Narratives

Across multi site interviews and workflow studies, clinicians describe template heavy, coding centered interfaces that privilege billing and standardized capture over nuanced narrative documenting of goals, preferences, and teach back confirmations [13,14]. Evidence suggests that limited or fragmented fields for concordance cues (e.g., goals of care rationales) make such information implicit or absent in notes, complicating retrieval and continuity across teams [21,22].

Theme B: Incomplete/Late Documentation of Adverse Effects, Pain, and SDOH

Rule based extraction and data mining analyses show pain as a frequent driver of unplanned care; however, symptom trajectories and supportive care triggers are often inconsistently recorded or delayed in the narrative, risking missed opportunities for early mitigation [9,10]. Palliative oncology notes reveal social distress and spiritual pain detectable in free text but absent in structured data, indicating hidden barriers to adherence when narrative signals are not systematically captured [1].

Theme C: Formal Tone and Stigmatizing Lexicon Undermine Trust and Comprehension

Lexicon driven studies identify stigmatizing language patterns in notes, with potential to transmit bias and alienate patients [12]. Mixed methods analyses in oncology point to benefits of plain language, person centred phrasing for understanding and anxiety reduction, while surveys of portal users indicate that documentation style functions as communication and shapes satisfaction and engagement [4,20,23].

Theme D: Time/Cognitive Burden Reduces Specificity and Explicit Concordance Recording

Audit log reviews document high navigation time, haphazard chart review patterns, and documentation fatigue that favor copy–paste over tailored narratives [14,16]. These pressures contribute to implicit recording of goals/values and fragmented adverse effect capture, reinforcing Theme A’s design constraints and Theme B’s coverage gaps [10,13].

3.6. Linkage to Research Questions

RQ1 (documentation patterns hindering concordance): Themes B and C show that under capture of adverse effects, pain, SDOH and formal orstigmatizing language systematically hinder retrieval of concordance relevant information and erode trust, respectively [1,4,9,10,12,20]. Theme A indicates that compliance driven designs reduce explicit documentation of goals, preferences, and decision rationales [13,14].

RQ2 (optimizing language and workflow): Evidence supports person centred prompts/fields, plain

language templates, stigma flagging, and audit log informed workflow redesign to expand narrative space, improve coverage/timing of key constructs, and enhance readability for patients accessing notes [13,14,20,21,23,24].

3.7. Guidance to Figures and Tables

Figure 1 provides the selection flow; Table ?? details corpus characteristics and methods; Table 4 summarizes theme prevalence across the EHR subset; and Figure 2 visualizes how design, coverage/timing, language/style, and workflow burden interact to shape concordance/adherence outcomes [1,4,9,10,12–14,16,20–24].

4. Discussion

4.1. Summary of Principal Findings

Across 23 studies of oncology documentation and EHR use, we identified four clinician-side themes tied to concordance and adherence: (1) EHR barriers and compliance-oriented design (template burden, coding-centric workflows), (2) incomplete/late documentation of adverse effects, pain, and social determinants of health, (3) formal note style and occasional stigmatizing language, and (4) time/cognitive burden limiting person-centred narratives [1,4,9,10,12–14,16,18,20]. These patterns align with informatics constraints (de-identification, corpus access) that have slowed oncology-specific NLP and sustained reliance on heterogeneous local corpora [5–8].

4.2. Interpretation in Relation to Communication, Trust, and Concordance

When adverse effects, pain, and SDOH are under-documented, teams miss timely opportunities to mitigate barriers to ongoing therapy [1,9,10]. Note style matters: formal, jargon-heavy, or stigmatizing phrasing can alienate patients, especially where portals expose notes directly [4,12,23]. Conversely, plain-language, stigma-free, and person-centred documentation supports understanding, trust, and participation—determinants consistently linked to adherence [29,30]. Studies classifying serious-illness communication show that goal-concordance cues are detectable in notes yet often remain implicit or absent, risking misalignment between care plans and patient priorities [21,22]. Overall, documentation quality is not merely archival; it is a communication act that shapes concordance.

4.3. Theoretical Framing: Communication, Trust, and Concordance

The findings of this review align with established theories of clinical communication and shared decision-making. According to relational communication theory, the quality of clinician–patient interactions including language tone, transparency, and responsiveness directly influences trust and engagement. When EHR documentation reflects formal, stigmatizing, or impersonal language, it may undermine these relational dynamics, particularly in oncology where patients often face complex, emotionally charged decisions.

Shared decision-making (SDM) frameworks emphasize the co-construction of care plans based on patient values, preferences, and goals. However, our synthesis shows that EHRs often lack structured fields or narrative space to document these elements explicitly. This gap risks misalignment between recorded care plans and actual patient priorities.

In this context, we define treatment concordance as the documented alignment between a patient's expressed goals, values, and preferences and the treatment plan recorded in the EHR. Concordance is not merely agreement on a clinical decision but includes the rationale, context, and shared understanding that underpin it. Effective concordance documentation requires both structured fields (e.g., goals-of-care checkboxes) and narrative entries that capture nuance, emotion, and evolving preferences.

Improving concordance documentation requires both technical solutions (e.g., NLP-ready templates) and a cultural shift toward communication-enhancing practices.

4.4. Implications for Nursing Practice, Documentation, and Service Design

Based on our findings, we recommend the following practical changes:

- Person-centred prompts and fields: add structured spaces for goals/values, decision rationale, and teach-back confirmation; pair with free-text for nuance [20,21].
- Adverse-effects/pain/SDOH capture: brief checklists with NLP-ready phrasing and prompts to reduce omission; trigger supportive care pathways (psycho-oncology, pain, social work) [1,9,10].
- Language guidance in notes: plain-language, inclusive, stigma-free templates with examples; automated flagging of stigmatizing terms for revision [4,12].
- Workflow relief: use audit-log insights to streamline navigation, reduce duplicative coding, and discourage copy–paste; reserve protected time for person-centred narrative entry [13,14,18].

Nursing leadership can champion documentation redesign as a safety and experience initiative, aligning with PCC and shared decision-making requirements and improving continuity across teams [20,29].

4.5. Practical Toolkit and Policy Alignment

To translate the findings of this review into actionable strategies, we propose a practical toolkit for improving EHR documentation in oncology. These recommendations align with the Australian Charter of Healthcare Rights and ACSQHC standards, especially in person-centred care, communication, and partnership.

Table 5. Toolkit for Enhancing Concordance-Focused EHR Documentation

Recommendation	Rationale and Policy Alignment
Add person-centred prompts and fields (e.g., goals, values, decision rationale)	Supports shared decision-making and aligns with the Charter’s principles of partnership and respect.
Use plain-language templates and examples	Enhances patient understanding and trust; aligns with ACSQHC communication standards.
Implement stigma-flagging tools (e.g., NLP-based detection of stigmatizing language)	Reduces bias and promotes dignity and respect in documentation.
Capture adverse effects, pain, and SDOH using structured and narrative fields	Improves early identification of barriers to adherence; supports safety and access principles.
Leverage audit-log data to streamline workflows	Reduces cognitive burden and documentation fatigue; aligns with safety and accountability standards.

Note: Recommendations are mapped to the Australian Charter of Healthcare Rights (2020) and ACSQHC documentation and communication standards.

4.6. Methodological Reflections and Limitations

EHR corpora vary widely by setting, EHR vendor, and policy; access is constrained by privacy and de-identification, complicating replication and external validity [5,6]. Text-mining pipelines differ in preprocessing, section segmentation, and lexicons; oncology-specific vocabularies for concordance and SDOH remain emergent [7,8]. Our discourse is an analytic synthesis is interpretive and cannot infer causality; observed documentation features may correlate with but not necessarily cause adherence outcomes. Finally, English language restriction and heterogeneity of clinical services may limit generalisability.

While this scoping review identifies key documentation patterns associated with treatment concordance and adherence, it does not establish causality. The generalizability of findings is further constrained by the heterogeneity of EHR systems, institutional practices, and the limited availability of oncology specific, de-identified corpora. Many included studies rely on local datasets or proprietary

systems, which may not reflect broader clinical contexts. Additionally, NLP methods varied widely in preprocessing, lexicon use, and classification strategies, with some pipelines lacking transparency or validation across settings. Such inconsistencies hinder cross study comparability and weaken the reliability of synthesized themes. Future research should prioritize standardized ontologies, shared corpora, and reproducible NLP workflows to enhance external validity and translational impact.

4.6.1. Gaps in the Literature and Future Directions

This review highlights several underexplored areas in the current literature. First, many included studies lacked detailed reporting on patient demographics, limiting insights into how documentation practices may differentially affect diverse populations. Few studies broke down findings by age, gender, ethnicity, or cancer type, despite known disparities in communication and adherence. Future research should prioritize inclusion of underrepresented groups and stratified analyses to ensure that documentation improvements benefit all patients equitably.

Second, there is a pressing need for standardized ontologies and shared vocabularies specific to oncology. Current NLP pipelines rely on heterogeneous lexicons, often adapted from general clinical contexts, which may not adequately capture oncology-specific constructs such as treatment concordance, adverse effect trajectories, or social determinants of health (SDOH). The development and adoption of open, validated ontologies for these domains would enhance the comparability, reproducibility, and clinical utility of NLP driven documentation analysis.

Closing these gaps is vital to developing equitable, scalable, and context sensitive documentation practices for person-centred oncology care.

NLP Techniques and Variability.

The included studies employed a range of NLP methods, including rule based extraction [10], machine learning classifiers [1], and n-gram or lexicon-based sentiment analysis [12,15]. However, few studies reported external validation or reproducibility metrics, and preprocessing steps (e.g., section segmentation, negation handling) were inconsistently described. This variability reduces comparability and may introduce bias.

Corpus Characteristics.

Table 6 summarizes corpus features across studies, including note types, corpus size, and language characteristics. Most studies relied on datasets from a single institution that are often with limited diversity in cancer types or patient demographics. The lack of standardized corpora and oncology specific annotation guidelines further constrains generalizability.

Table 6. Summary of corpus characteristics in included studies.

Study	Corpus Size	Note Types	Language Features
Masukawa et al. (2022) [1]	1,200	Palliative care notes	Japanese; spiritual pain terms
Tamang et al. (2015) [10]	5,000+	Progress notes	Pain-related expressions
Himmelstein et al. (2022) [12]	10,000+	Admission notes	Stigma lexicon, U.S. English
Rahimian et al. (2019) [15]	3,000	Oncology notes (OpenNotes)	n-gram frequency, formal tone
Heckemann et al. (2022) [20]	500	Discharge letters	Person-centred phrasing

Note: Corpus sizes and note types vary widely across studies. Language features reflect the primary focus of each study’s NLP or linguistic analysis.

Bias and Validation.

Several studies lacked transparency regarding model training data, annotation protocols, or inter-rater reliability. This raises concerns about potential bias, particularly in detecting stigmatizing language or adverse effects, and limits confidence in generalizing findings. Future NLP research in oncology should focus on open-source corpora, multi-site validation, and standardized evaluation metrics to strengthen methodological rigor.

4.7. Future Research

We highlight four priorities:

1. Oncology-specific documentation lexicons for concordance, preference-sensitive decisions, adverse-effect trajectories, and SDOH to improve indicator detection across sites [7,8].
2. Prospective trials of note style interventions (plain language prompts, stigma flagging, concordance fields) with outcomes in understanding, trust, and adherence [4,12,29].
3. Audit log informed workflow redesign evaluated for time burden, documentation completeness, and downstream unplanned care [10,13,14].
4. Patient portal co-design to pair readable notes with tailored education and teach back summaries, closing loops between documentation and patient comprehension [23,24].

5. Conclusions

Documentation quality in oncology is a modifiable determinant of communication, concordance, and adherence. This review identified four clinician-side barriers design constraints, incomplete documentation, formal or stigmatizing language, and workflow burden that limit the recording and retrieval of person-centred information.

By addressing these barriers through person-centred prompts, stigma-free language, improved capture of adverse effects and SDOH, and workflow redesign, nursing and service leaders can enhance trust, engagement, and continuity of care. These strategies directly support the principles of shared decision-making and align with national safety and quality standards.

Our findings respond to both research questions: they identify documentation patterns that hinder concordance and propose actionable design and language interventions to improve adherence outcomes. Future research should focus on scalable implementation, equity in documentation practices, and co-design with patients.

Author Contributions: Conceptualization, L.W.; methodology, L.W., R.G. and X.Z; formal analysis, L.W.; investigation, L.W; writing—original draft preparation, L.W; writing—review and editing, L.W., R.G. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any funding.

Institutional Review Board Statement: Ethical review was not needed for this study.

Public Involvement Statement: There was no public involvement in any aspect of this research.

Guidelines and Standards Statement: This manuscript was drafted against the PRISMA-ScR checklist for reporting of scoping reviews.

Use of Artificial Intelligence: Generative AI tools were not used in drafting this manuscript. However, the study relies on AI-based analytical methods such as Natural Language Processing (NLP) and sentiment analysis to examine patient-generated narratives. Existing AI tools were employed strictly for data analysis, not for content creation.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Masukawa, K.; Aoyama, M.; Yokota, S.; Nakamura, J.; Ishida, R.; Nakayama, M.; Miyashita, M. Machine learning models to detect social distress, spiritual pain, and severe physical psychological symptoms in terminally ill patients with cancer from unstructured text data in electronic medical records. *Palliative Medicine* **2022**, *36*(8), 1207–1216. <https://doi.org/10.1177/02692163221105595>.
2. Percha, B. Modern Clinical Text Mining: A Guide and Review. *Annual Review of Biomedical Data Science* **2021**, *4*(1), 165–187. <https://doi.org/10.1146/annurev-biodatasci-030421-030931>.
3. Sikorskii, A.; Wyatt, G.; Tamkus, D.; Victorson, D.; Rahbar, M. H.; Ahn, S. Concordance between patient reports of cancer-related symptoms and medical records documentation. *J Pain Symptom Manage* **2012**, *44*(3), 362–372. <https://doi.org/10.1016/j.jpainsymman.2011.09.017>.
4. Alpert, J. M.; Morris, B. B.; Thomson, M. D.; Matin, K.; Sabo, R. T.; Brown, R. F. Patient access to clinical notes in oncology: A mixed method analysis of oncologists' attitudes and linguistic characteristics towards notes. *Patient Educ Couns* **2019**, *102*(10), 1917–1924. <https://doi.org/10.1016/j.pec.2019.05.008>.
5. Chapman, W.; Nadkarni, P.; Hirschman, L.; D'Avolio, L.; Savova, G.; Uzuner, O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. *Journal of the American Medical Informatics Association: JAMIA* **2011**, 540–543. <https://doi.org/10.1136/amiajnl-2011-000465>.
6. Johnson, A.; Pollard, T.; Horng, S.; Anthony, L.; Mark, R. MIMIC-IV-Note: Deidentified free-text clinical notes. *PhysioNet* **2023**, Version 2.2. <https://doi.org/10.13026/1n74-ne17>.
7. Zeng, J.; Banerjee, I.; Henry, A. S.; Wood, D. J. Natural Language Processing to Identify Cancer Treatments With Electronic Medical Records. *JCO: Clinical Cancer Informatics* **2021**, *5*. <https://doi.org/10.1200/CCI.20.0173>.
8. Gholipour, M.; Khajouei, R.; Amiri, P.; Hajesmaeel Gohari, S.; Ahmadian, L. Extracting cancer concepts from clinical notes using natural language processing: a systematic review. *BMC Bioinformatics* **2023**, *24*. <https://doi.org/10.1186/s12859-023-05480-0>.
9. Kuo, C. C.; Wang, H. H.; Tseng, L. P. Using data mining technology to predict medication-taking behaviour in women with breast cancer: A retrospective study. *Nurs Open* **2022**, *9*(6), 2646–2656. <https://doi.org/10.1002/nop2.963>.
10. Tamang, S.; Patel, M.; Blayney, D.; Kuznetsov, J.; Finlayson, S.; Vetteth, Y. Detecting unplanned care from clinician notes in electronic health records. *J Oncol Pract* **2015**, *11*, e313–e319. <https://doi.org/10.1200/JOP.2014.002741>.
11. Cho, J. E.; Tang, N.; Pitaro, N.; Bai, H.; Cooke, P. V.; Arvind, V. Sentiment Analysis of Online Patient-Written Reviews of Vascular Surgeons. *Ann Vasc Surg* **2023**, *88*, 249–255. <https://doi.org/10.1016/j.avsg.2022.07.016>.
12. Himmelstein, G.; Bates, D.; Zhou, L. Examination of Stigmatizing Language in the Electronic Health Record. *JAMA Network Open* **2022**, *5*(1), e2144967. <https://doi.org/10.1001/jamanetworkopen.2021.44967>.
13. Patel, M. R.; Friese, C. R.; Mendelsohn-Victor, K.; Fauer, A. J.; Ghosh, B.; Bedard, L.; Griggs, J. J.; Manojlovich, M. Clinician Perspectives on Electronic Health Records, Communication, and Patient Safety Across Diverse Medical Oncology Practices. *Journal of Oncology Practice* **2019**, *15*(6), e529–e536. <https://doi.org/10.1200/JOP.18.00507>.
14. Huilgol, Y. S.; Adler-Milstein, J.; Ivey, S. L.; Hong, J. C. Opportunities to use electronic health record audit logs to improve cancer care. *Cancer Medicine* **2022**, *11*(17), 3296–3303. <https://doi.org/10.1002/cam4.4690>.
15. Rahimian, M.; Warner, J.; Jain, S.; Davis, R.; Zerillo, J.; Joyce, R. Significant and Distinctive n-Grams in Oncology Notes: A Text-Mining Method to Analyze the Effect of OpenNotes on Clinical Documentation. *JCO Clin Cancer Inform* **2019**, *3*, 1–9. <https://doi.org/10.1200/CCI.19.00012>.
16. Warner, J.; Hochberg, E. Where is the EHR in oncology? *J Natl Compr Canc Netw* **2012**, *10*(5), 584–588. <https://doi.org/10.6004/jnccn.2012.0060>.
17. Asan, O.; Nattinger, A. B.; Gurses, A. P.; Tyszka, J. T.; Yen, T. W. F. Oncologists' Views Regarding the Role of Electronic Health Records in Care Coordination. *JCO Clinical Cancer Informatics* **2018**, *2*, 1–12. <https://doi.org/10.1200/CCI.17.00118>.
18. Burke, H. B. Standardized Documentation Is Not the Solution to Reduce Physician Time in the Electronic Health Record. *JAMA Oncology* **2023**, *9*(8), 1151–1152. <https://doi.org/10.1001/jamaoncol.2023.1523>.
19. Geskey, J. M.; Kodish-Wachs, J.; Blonsky, H.; Hohman, S. F.; Meurer, S. National Documentation and Coding Practices of Noncompliance: The Importance of Social Determinants of Health and the Stigma of African-American Bias. *Am J Med Qual* **2023**, *38*(2), 87–92. <https://doi.org/10.1097/jmq.0000000000000112>.

20. Heckemann, B.; Chaaya, M.; Jakobsson Ung, E.; Olsson, D. S.; Jakobsson, S. Finding the Person in Electronic Health Records. A Mixed-Methods Analysis of Person-Centered Content and Language. *Health Communication* **2022**, *37*(4), 418–424. <https://doi.org/10.1080/10410236.2020.1846275>.
21. Davoudi, A.; Tissot, H.; Doucette, A.; et al. Using Natural Language Processing to Classify Serious Illness Communication with Oncology Patients. *AMIA Jt Summits Transl Sci Proc* **2022**, 168–177. PMID: [PMC9285137](https://pubmed.ncbi.nlm.nih.gov/39285137/)
22. Geerse, O. P.; Lamas, D. J.; Bernacki, R. E.; Sanders, J. J.; Paladino, J.; Berendsen, A. J.; Hiltermann, T. J. N.; Lindvall, C.; Fromme, E. K.; Block, S. D. Adherence and Concordance between Serious Illness Care Planning Conversations and Oncology Clinician Documentation among Patients with Advanced Cancer. *J Palliat Med* **2021**, *24*(1), 53–62. <https://doi.org/10.1089/jpm.2019.0615>.
23. Rexhepi, H.; Moll, J.; Huvila, I. Online electronic healthcare records: Comparing the views of cancer patients and others. *Health Informatics Journal* **2020**, *26*(4), 2915–2929. <https://doi.org/10.1177/1460458220944727>.
24. Griffin, J. M.; Kroner, B. L.; Wong, S. L.; Preiss, L.; Wilder Smith, A.; Cheville, A. L.; Mitchell, S. A.; Lancki, N.; Hassett, M. J.; Schrag, D.; Osarogiagbon, R. U.; Ridgeway, J. L.; Cella, D.; Jensen, R. E.; Flores, A. M.; Austin, J. D.; Yanez, B. Disparities in electronic health record portal access and use among patients with cancer. *J Natl Cancer Inst* **2024**, *116*(3), 476–484. <https://doi.org/10.1093/jnci/djad225>.
25. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. <https://doi.org/10.7326/M18-0850>.
26. Moralejo, D.; Ogunremi, T.; Dunn, K. Critical Appraisal Toolkit (CAT) for assessing multiple types of evidence. *Can Commun Dis Rep* **2017**, *43*(9), 176–181. <https://doi.org/10.14745/ccdr.v43i09a02>.
27. Skeat, J.; Roddam, H. The qual-CAT:
Applying a rapid review approach to qualitative research to support clinical decision-making in speech-language pathology practice. *Evidence-Based Communication Assessment and Intervention* **2019**, *13*(1-2), 3–14. <https://doi.org/10.1080/17489539.2019.1600292>.
28. CASP. Critical Appraisal Skills Programme. *CASP UK - OAP Ltd* **2023**. Retrieved 24.09.23. <https://casp-uk.net/casp-tools-checklists/>.
29. Hillen, M. A.; de Haes, H. C. J. M.; Stalpers, L. J. A.; Klinkenbijn, J. H. G.; Eddes, E. H.; Butow, P. N.; van der Vloodt, J.; van Laarhoven, H. W. M.; Smets, E. M. A. How can communication by oncologists enhance patients' trust? An experimental study. *Annals of Oncology* **2014**, *25*(4), 896–901. <https://doi.org/10.1093/annonc/mdl027>.
30. Zolnieriek, K. B.; Dimatteo, M. R. Physician communication and patient adherence to treatment: a meta-analysis. *Med Care* **2009**, *47*(8), 826–834. <https://doi.org/10.1097/MLR.0b013e31819a5acc>.
31. Australian Commission on Safety and Quality in Health Care. Australian Charter of Healthcare Rights (Second Edition). *ACSQHC* **2020**. <https://www.safetyandquality.gov.au/publications-and-resources/resource-library/australian-charter-healthcare-rights-second-edition-a4-accessible>.
32. Ahmad, B.; Jun, S. Sentiment Analysis of Cancer Patients About Their Treatment During the Peak Time of Pandemic COVID-19. In *2021 4th International Conference on Computing and Information Sciences (ICCIS)* **2021**, 29–30 Nov. <https://doi.org/10.1109/ICCIS54243.2021.9676393>
33. Amossy, R. Argumentation in Discourse: A Socio-discursive Approach to Arguments. *Informal Logic* **2009**, *29*, 252–267. <https://doi.org/10.22329/il.v29i3.2843>.
34. An, Y.; Fang, Q.; Wang, L. Enhancing patient education in cancer care: Intelligent cancer patient education model for effective communication. *Computers in Biology and Medicine* **2024**, *169*, 107874. <https://doi.org/10.1016/j.compbiomed.2023.107874>.
35. Arditi, D.; Gilles, I.; Lesage, S.; Griesser, A.-C.; Bienvenu, C.; et al. Computer-assisted textual analysis of free-text comments in the Swiss Cancer Patient Experiences (SCAPE) survey. *BMC Health Serv Res* **2020**, *20*, 1029. <https://doi.org/10.1186/s12913-020-05873-4>.
36. Simoulin, A.; Thiebaut, N.; Neuberger, K.; Ibnouhsein, I.; Brunel, N.; Viné, R.; Bousquet, N.; Latapy, J.; Reix, N.; Molière, S.; Lodi, M.; Mathelin, C. From free-text electronic health records to structured cohorts: Onconum, an innovative methodology for real-world data mining in breast cancer. *Computer Methods and Programs in Biomedicine* **2023**, *240*, 107693. <https://doi.org/10.1016/j.cmpb.2023.107693>.
37. Xin, J.; Hjaltelin, J.X.; Novitski, S.I.; Jørgensen, I.F.; Siggaard, T.; Vulpius, S.A.; Westergaard, D.; Johansen, J.S.; Chen, I.M.; Jensen, L.J.; Brunak, S. Pancreatic cancer symptom trajectories from Danish registry data and free text in electronic health records. *eLife* **2023**, *12*:e84919. <https://doi.org/10.7554/eLife.84919>.

38. Elbers, D.C.; La, J.; Minot, J.R.; Gramling, R.; Brophy, M.T.; Do, N.V.; Fillmore, N.R.; Dodds, P.S.; Danforth, C.M. Sentiment analysis of medical record notes for lung cancer patients at the Department of Veterans Affairs. *PLOS ONE* **2023**, *18*(1), e0280931. <https://doi.org/10.1371/journal.pone.0280931>.
39. Leis, A.; Casadevall, D.; Albanell, J.; Posso, M.; Macià, F.; Castells, X.; Ramírez-Anguaita, J.M.; Martínez Roldán, J.; Furlong, L.I.; Sanz, F.; Ronzano, F.; Mayer, M.A. Exploring the Association of Cancer and Depression in Electronic Health Records: Combining Encoded Diagnosis and Mining Free-Text Clinical Notes. *JMIR Cancer* **2022**, *8*(3), e39003. <https://doi.org/10.2196/39003>.
40. Mashima, Y.; Tamura, T.; Kunikata, J.; Tada, S.; Yamada, A.; Tanigawa, M.; Hayakawa, A.; Tanabe, H.; Yokoi, H. Using Natural Language Processing Techniques to Detect Adverse Events From Progress Notes Due to Chemotherapy. *Cancer Informatics* **2022**, *21*, 11769351221085064. <https://doi.org/10.1177/11769351221085064>.
41. Jensen, K.; Soguero-Ruiz, C.; Mikalsen, K.Ø.; Lindsetmo, R.-O.; Kouskoumvekaki, I.; Girolami, M.; Skrovseth, S.O.; Augestad, K.M. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Scientific Reports* **2017**, *7*, 46226. <https://doi.org/10.1038/srep46226>.
42. Feldman, K.; Hazekamp, N.; Chawla, N.V. Mining the Clinical Narrative: All Text Are Not Equal. *2016 IEEE International Conference on Healthcare Informatics (ICHI)* **2016**, 306–314. <https://doi.org/10.1109/ICHI.2016.37>.
43. Ganesan, K.; Lloyd, S.; Sarkar, V. Discovering Related Clinical Concepts Using Large Amounts of Clinical Notes. *Biomedical Engineering and Computational Biology* **2016**, *7*(Suppl 2), 27–33. <https://doi.org/10.4137/BECB.S36155>.
44. Falotico, R.; Liberati, C.; Zappa, P. Identifying Oncological Patient Information Needs to Improve e-Health Communication: a preliminary text-mining analysis. *Quality and Reliability Engineering International* **2015**, *31*(7), 1115–1126. <https://doi.org/10.1002/qre.1853>.
45. Hamid, M.S.; Valicevic, A.; Brenneman, B.; Niziol, L.M.; Stein, J.D.; Newman-Casey, P.A. Text Parsing-Based Identification of Patients with Poor Glaucoma Medication Adherence in the Electronic Health Record. *American Journal of Ophthalmology* **2020**, *222*, 54–59. <https://doi.org/10.1016/j.ajo.2020.09.008>.
46. Zhu, V.J.; Lenert, L.A.; Bunnell, B.E.; Obeid, J.S.; Jefferson, M.; Hughes Halbert, C. Automatically identifying social isolation from clinical narratives for patients with prostate cancer. *BMC Medical Informatics and Decision Making* **2019**, *19*, 43. <https://doi.org/10.1186/s12911-019-0795-y>.
47. Datta, S.; Bernstam, E.V.; Roberts, K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *Journal of Biomedical Informatics* **2019**, *100*, 103301. <https://doi.org/10.1016/j.jbi.2019.103301>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.