Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Utterance-level operational definitions of shared decision-making for primary care: an empirical annotation study of physician-patient conversations

Eunsuk Chang
Republic of Korea Air Force Aerospace Medicine Research Center
Shinwha Ro
Independent researcher
Sang-Ho Yoo
karmaboy@hanyang.ac.kr

Hanyang University

Research Article

Keywords: Shared decision-making, Primary health care, Physician-patient relations, Annotation, Natural language processing, Artificial intelligence

Posted Date: June 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9393787/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Shared decision-making (SDM) in primary care often occurs through multiple brief, overlapping, and partially explicit decisions. Frameworks developed in specialty settings may therefore fail to capture how SDM is enacted in routine primary care, limiting both human evaluation and the development of automated assessment tools. We aimed to derive empirically grounded, utterance-level operational definitions of SDM from primary care dialogue.

Methods

We conducted a qualitative annotation and framework refinement study using 405 de-identified physician–patient transcripts from the Content Coding for Contextualization of Care dataset. Starting with the Makoul-Clayman model, we developed a literature-informed initial codebook for the nine SDM elements and applied it at the utterance level. Two annotators independently coded the transcripts, resolved disagreements through iterative consensus with adjudication by a third reviewer, and revised the definitions until the codebook stabilized. After stabilization, the final schema was applied across the corpus and summarized descriptively.

Results

The original framework remained a useful conceptual scaffold. However, several elements required operational refinement in primary care. What is traditionally coded as patient values/preferences most often appeared as brief, situated rationales for accepting or resisting an option, leading to relabeling as “Reason for choice.” The broad category of doctor knowledge/recommendations was operationalized as two separate annotatable codes—“Information provision” and “Clinician recommendation”—because these functions were interactionally distinct. Across the corpus, SDM elements were unevenly distributed, implicit, and frequently embedded within rapid topic shifts. Decision closure and clinician recommendations were common, whereas explicit checks for understanding and discussions of patient ability/self-efficacy were rare.

Conclusions

SDM in routine primary care is less linear and more context-dependent than many existing models. The revised utterance-level definitions presented here offer a more precise representation of SDM in primary care, support more context-sensitive evaluation, and provide an annotation target for future AI-based assessments of clinical dialogue.

Background

Shared decision-making (SDM) is a collaborative process in which clinicians and patients use the best available evidence to choose among reasonable options in light of what matters to the patient [1]. It has been associated with better patient knowledge, lower decisional conflict, and more value-concordant care; however, its uptake in routine practice remains uneven [2–5]. Recent policy and professional initiatives continue to call for the wider implementation of SDM [6]. One persistent obstacle is measurement; determining whether SDM occurs requires close analysis of clinical encounters, and manual coding is difficult to scale [7, 8].

This measurement problem is particularly acute in primary care settings. Unlike many specialty consultations organized around a single high-salience choice, primary care visits often involve several smaller decisions—medication changes, preventive services, diagnostic testing, or self-management steps—within a single conversation [9–12]. These decisions can be brief, partially implicit, or distributed over time. Frameworks and coding systems developed in more bounded decision contexts therefore risk mis-specifying what SDM looks like in everyday primary care [7, 13, 14].

This issue is increasingly important for health informatics. Recent work suggests that large language models (LLMs) can approximate human scoring on established SDM measures in selected settings [15], and broader reviews have shown growing interest in using LLMs for clinical documentation, communication, and decision support [16]. However, an automated assessment can only be valid for the categories it is designed to detect. If the underlying coding framework does not reflect the interactional texture of primary care, automation will reproduce conceptual errors at scale.

Therefore, we aimed to derive operational definitions of SDM elements that are usable at the utterance level and grounded in real primary care dialogues. Using a corpus of physician–patient transcripts, we iteratively refined the Makoul-Clayman SDM framework [17] to produce a primary care–adapted annotation schema suitable for both human analysis and future AI-based assessment.

Methods

Study design and data source

This was a qualitative annotation and framework refinement study based on a secondary analysis of physician–patient transcripts. This study was designed to translate a theory-based SDM model into observable conversational units that can be applied consistently by human coders and, eventually, by computational systems.

We used the Content Coding for Contextualization of Care (4C) dataset, a public collection of 405 de-identified outpatient primary care transcripts from the U.S. Veterans Health Administration (VHA) [18, 19]. The corpus captures routine visits rather than encounters staged around a single focal decision, making it well suited to examining how SDM is expressed under ordinary primary care conditions. The

source data were de-identified before being publicly released. As the present study involved a secondary analysis of publicly available, de-identified transcripts without contact with participants, an additional ethics review was not required under institutional policy.

Transcript length varied substantially across the corpus; Supplementary Figure S1 summarizes the distributions of token counts and utterance counts. Because one utterance can receive more than one code, the element frequencies reported below represent coded instances rather than visits or discrete decision episodes. No inferential statistics were planned, as the purpose of the study was definitional refinement rather than hypothesis testing.

Coding framework and initial codebook

We used the Makoul-Clayman integrative model as the conceptual starting point, as it remains one of the most widely used syntheses of core SDM elements [17]. To make the model usable at the utterance level, we drafted an initial codebook based on the original framework [17], its subsequent operationalization by Clayman et al. [14], comparative work on SDM coding systems [13], and decision support measures such as DSAT and OPTION [20, 21].

The initial codebook retained the nine-element scaffold of the original framework: define/explain the problem; present options; discuss pros/cons; patient values/preferences; discuss patient ability/self-efficacy; doctor knowledge/recommendations; check/clarify understanding; make or explicitly defer decisions; and arrange follow-ups. The initial definitions in the literature are summarized in Table 1.

Table 1
Literature-informed initial operational definitions for shared decision-making elements.

Element	Operational definition / coding guidance
Define/explain problem	Introduces or frames an issue as requiring clinical attention or a decision. May signal uncertainty, urgency, or the need to choose among reasonable paths.
Present options	States one or more possible management approaches, including no action when relevant. Options may be explicit or implicit.
Discuss pros/cons	Describes potential benefits, harms, burdens, or consequences of options, including the consequences of inaction.
Patient values/preferences	Elicits or explores what the patient prefers, values, fears, or hopes for in relation to the decision.
Discuss patient ability/self-efficacy	Assesses the patient's ability, readiness, confidence, or practical capacity to carry out a chosen option.
Doctor knowledge/recommendations	Provides decision-relevant clinical information and/or expresses a preferred course of action based on clinical judgment.
Check/clarify understanding	Checks whether the patient understands decision-relevant information or invites clarification beyond simple confirmation.
Make or explicitly defer decision	Marks agreement, acceptance, rejection, or explicit postponement of a decision.
Arrange follow-up	Specifies monitoring, reassessment, or next steps directly linked to the decision.

Annotation procedure and iterative refinement

The unit of analysis was the utterance, defined as a complete speaking turn ending in standard punctuation. Utterance-level coding was selected because it preserves the local sequencing of decision talk and is compatible with the subsequent computational modeling of dialogue.

Two annotators (EC and SR) independently coded the transcripts, and disagreements were resolved by a third reviewer (SHY). Each utterance could receive up to two SDM codes when multiple functions were performed. When the same function extended across consecutive utterances, each utterance was coded separately to preserve sequential detail. Figure 1 shows an illustrative example of utterance-level multi-label coding.

Codebook refinement was iterative. The annotators met regularly to compare interpretations, identify boundary problems, and revise the definitions when recurring ambiguities emerged. The framework was considered stable when successive rounds of coding no longer produced substantive changes to the operational definitions.

Because the early rounds were devoted to developing, rather than testing, a fixed codebook, formal interrater coefficients were not estimated at the outset. After the definitions were stabilized, a subset of transcripts was independently coded to confirm that the finalized codebook could be applied consistently. We report this as a framework development study and treat the resulting frequencies as descriptive.

The final schema preserved the original conceptual architecture but operationalized one element, doctor knowledge/recommendations, as two distinct annotatable codes: information provision and clinician recommendation. Therefore, the revised schema remained anchored in the nine-element model, yielding ten observable codes for annotation. The final definitions are presented in Table 2.

Results

Refinement of SDM element definitions

Across the 405 transcripts, the Makoul-Clayman framework remained recognizable but required targeted revision at several recurrent fault lines. Three adjustments were especially important: re-specifying patient values/preferences as “Reason for choice”; splitting doctor knowledge/recommendations into “Information provision” and “Clinician recommendation”; and clarifying that both problem definition and decision closure were often signaled implicitly rather than through overt framing language.

Relabeling patient values/preferences reflected a consistent empirical pattern. In this corpus, patients seldom articulated values in abstract terms; instead, they offered concise, situated rationales for accepting, declining, or resisting an option, such as concerns about burden, reluctance to add medication, cost, inconvenience, or prior experience. The revised label “Reason for choice” more accurately captured what was observable in the talk.

The original category of doctor knowledge/recommendations also proved too coarse for primary care dialogue. Informational explanations and treatment endorsements were often adjacent but not interactionally equivalent. Information provision supplied facts or interpretations relevant to the decision, whereas clinician recommendations proposed a preferred course of action and oriented the conversation toward uptake, negotiation, or refusal. Treating them as separate codes improves both conceptual clarity and annotation precision.

Table 2
Empirically refined operational definitions of shared decision-making elements for primary care.

Element	Operational definition / coding guidance
1. Define/explain problem	Introduces or frames an issue as requiring clinical attention or a decision. Includes explicit problem naming and conversational cues that open or shift a decision topic. May signal urgency or uncertainty. Excludes general information unrelated to a decision.
2. Present options	States one or more plausible management paths, including watchful waiting or no action when relevant. Options may be contrasted explicitly or implied by offering one actionable path in a context where alternatives are available. Excludes mere endorsement of a preferred plan without signaling choice.
3. Discuss pros/cons	Describes anticipated benefits, harms, burdens, or consequences of options, including the consequences of inaction. Focuses on prospective implications rather than retrospective appraisal. Excludes disease education not linked to choosing among options.
4. Reason for choice	Captures the patient's stated rationale for accepting, declining, or preferring an option. Typically appears as brief, local justification rooted in concerns, circumstances, or goals rather than abstract values language. Excludes general preferences not tied to a current decision.
5. Discuss patient ability/self-efficacy	Addresses whether the patient can realistically carry out a plan, including willingness, confidence, cost, logistics, and anticipated adherence. Excludes general counseling not tied to implementing a specific option.
6a. Information provision	Provides decision-relevant factual or explanatory information about the condition, the options, or likely outcomes. Restricted to information that could reasonably inform the immediate decision. Excludes tangential education and administrative detail.
6b. Clinician recommendation	Expresses the clinician's preferred option before the decision is finalized. The recommendation remains open to patient uptake, modification, or refusal. Excludes final agreement or plan confirmation.
7. Check/clarify understanding	Actively tests, elicits, or repairs the patient's understanding of decision-relevant information. Includes requests for explanation, elaboration, or questions that invite demonstration of comprehension. Excludes perfunctory yes/no checks that do not probe understanding.
8. Make or explicitly defer decision	Marks acceptance, rejection, confirmation, or deliberate postponement of a decision. Includes brief acknowledgements when their function as decision closure is clear from surrounding talk. Excludes recommendations that have not yet been taken up.
9. Arrange follow-up	Specifies what will happen next because of the decision, such as monitoring, reassessment, scheduling, testing, or implementation steps. Excludes generic administrative closure unrelated to the decision.

Frequency and distribution of coded elements

After codebook stabilization, 9,781 SDM-element instances were identified across the corpus (Fig. 2). The most frequent codes were making or explicitly deferring decisions (n = 3,985) and clinician recommendations (n = 2,926), followed by presenting options (n = 1,012) and arranging follow-up (n =

514). Defining/explaining problems (n = 475), reasons for choice (n = 313), information provision (n = 188), and discussion of pros/cons (n = 187). In contrast, discussions of patient ability/self-efficacy (n = 95) and checking/clarifying understanding (n = 86) were less common. These counts describe the observable verbalized behavior in this dataset; they should not be interpreted as a hierarchy of conceptual importance.

Interactional characteristics of primary care SDM

Decision-making in primary care is typically brief, topic-shifting, and uneven. A single visit often contains several decision threads, such as medication adjustment, preventive screening, and follow-up planning, each expressed through a different subset of SDM elements. Rather than unfolding as a single linear sequence, SDM appears as a set of short, local episodes embedded within a larger visit.

Option presentation and clinician recommendations frequently co-occur, reflecting a pragmatic primary care style in which a clinician offers a preferred action while leaving room for patient response. In contrast, self-efficacy assessment and explicit checking of understanding are seldom verbalized. Coding also depends heavily on the local context. The same lexical material can function differently depending on what has just occurred in the conversation, and brief acknowledgments can count as decision closures only when they clearly ratify a preceding recommendation or option.

Taken together, the revised schema captures SDM in primary care as less ceremonious and more interactionally compressed than many formal models imply. The primary care signal is not the orderly presence of every element in every decision, but the patterned way in which choice, recommendation, justification, and follow-up are woven into routine talk.

Discussion

Our central contribution is as much methodological as it is conceptual. Before SDM can be measured reliably or automated credibly in primary care, it must be defined at the level at which primary care conversations actually unfold. Using a large corpus of routine visits, we show that the Makoul-Clayman model remains a useful scaffold [17]. Still, its operational boundaries require adjustment when applied to brief, multi-issue encounters.

These refinements were sequential. First, what classic SDM frameworks describe as patient values/preferences is usually observable here as a local rationale for accepting or resisting an option, rather than as extended deliberation about abstract values. Second, the combined category of doctor knowledge/recommendations concealed two distinct clinician actions: explanations and endorsements. Separating these actions is important analytically because they invite different types of patient responses and likely pose different challenges for automated detection.

These findings have implications for SDM measurement. Encounter-level scores may flatten meaningful variation when one visit involves several small decisions, each with a different communicative profile. Therefore, our results support more topic- or decision-level assessments, especially in primary care, where the conversational unit of interest is often smaller than the visit itself. They also suggest that instruments requiring overt verbalization of all SDM elements may underestimate decision-sharing when important work is accomplished implicitly or across adjacent visits [7, 8, 11, 12].

At the same time, the rarity of self-efficacy talk and explicit understanding checks should not be normalized away. Earlier SDM coding also treated self-efficacy as a distinct element of decision support [14], and self-efficacy was clinically consequential in primary care populations with multimorbidity [22]. Decision aids may help patients compare options and clarify what matters to them [23], but they do not eliminate the need for clinicians to assess understanding and feasibility during the encounter. Clinicians should also not assume that deference implies disengagement; some patients deliberately delegate decisions because of trust rather than lack of interest [24].

These findings are directly relevant to AI-based assessments. Recent studies have indicated that LLMs can approximate human ratings of observer-based SDM measures in selected specialty contexts [15]. Our study suggests that primary care requires more than utterance-by-utterance classification of isolated sentences. Automated systems must track topic shifts, speaker roles, sequential uptake, and the contextual status of seemingly minimal responses. In this sense, the present codebook is not only a descriptive framework but also an annotation target designed for computational use.

For implementation, AI support should aim to surface missing conversational work rather than merely produce a global SDM score. A clinically useful system might identify when options are presented without any discussion of consequences, or when a recommendation is accepted without checking understanding or feasibility. Such targeted feedback is more actionable for clinicians and better aligned with the fragmented ecology of primary care than a single summary score [25].

This study has several limitations. The corpus came from a single health system and reflects routine U.S. VHA primary care; therefore, its transferability to other settings, populations, and languages remains uncertain [18, 19]. The transcripts captured speech but not gaze, gestures, or prosody, all of which may contribute to decision-making. Formal inter-rater reliability statistics were not calculated during codebook development, and future studies should quantify the reliability of the final schema using independent samples. Finally, because the analysis was based on single-visit transcripts, it could not fully capture decisions distributed across multiple encounters.

Despite these limitations, this study identifies a practical route between two unsatisfactory options: applying specialty-derived SDM categories unchanged to primary care or abandoning structured assessment altogether. A primary care–adapted operational schema allows the field to preserve its conceptual rigor while respecting the actual cadence of routine visits.

Conclusions

This study develops utterance-level operational definitions of SDM for primary care by refining an established model against real physician–patient conversations. The resulting schema makes primary care decision-making more legible by distinguishing information from recommendations, capturing patient reasoning as situated justification, and recognizing the importance of implicit closure and follow-up.

These refinements support more credible human evaluations and provide a stronger foundation for AI systems designed to assess and support SDM in daily clinical care.

Abbreviations

4C, Content Coding for Contextualization of Care

AI, artificial intelligence

DSAT, Decision Support Analysis Tool

LLM, large language model

SDM, shared decision-making

VHA, Veterans Health Administration

Declarations

Ethics approval and consent to participate

The source dataset was de-identified before public release. The present study involved secondary analysis of publicly available de-identified transcripts and no contact with participants; accordingly, additional ethics review and informed consent for this secondary analysis were not required under institutional policy.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Korea Health Technology R&D Project through the Patient-Doctor Shared Decision-Making Research Center (grant HI20C1234), Ministry of Health & Welfare, Republic of Korea. The funder had no role in study design, analysis, interpretation of the data, or manuscript preparation.

Author Contribution

EC conceived the study, curated the data, performed the formal analysis, conducted coding, created the visualizations, and drafted the manuscript. SR contributed to data curation, coding, methodology, project administration, and validation. SHY conceived and supervised the study, contributed to analysis and methodology, secured funding and resources, and critically revised the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors thank the U.S. Veterans Health Administration for making the 4C conversation dataset publicly available and acknowledge the patients and clinicians whose recorded visits made this work possible.

Data Availability

The transcripts analyzed during the current study are publicly available from the U.S. Department of Veterans Affairs Open Data Portal: https://www.data.va.gov/dataset/Physician-patient-transcripts-with-4C-coding-analy/4qbs-wgct. The literature-informed initial codebook, the final operational definitions, and aggregated descriptive outputs are included in this article and its supplementary material. Study-specific derived annotation files are not publicly available, as they are intended for use in training an AI application the authors plan to develop for clinical shared decision-making. However, they may be obtained from the corresponding author upon reasonable request, subject to the terms of use of the source dataset.

References

1. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med. 2012;27:1361–7.
2. Stacey D, Lewis KB, Smith M, Carley M, Volk RJ, Douglas EE, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2024;1:CD001431.
3. Waddell A, Lennox A, Spassova G, Bragge P. Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. Implement Sci. 2021;16:74.
4. Jaeken J, Billiouw C, Mertens L, Van Bostraeten P, Bekkering G, Vermandere M, et al. A systematic review of shared decision making training programs for general practitioners. BMC Med Educ.

2024;24:592.

5. Holmes-Rovner M, Valade D, Orlowski C, Draus C, Nabozny-Valerio B, Keiser S. Implementing shared decision-making in routine practice: barriers and opportunities. Health Expect. 2000;3:182–91.
6. Perpetua EM, Palmer R, Le VT, Al-Khatib SM, Beavers CJ, Beckman JA, et al. JACC: advances expert panel perspective: Shared decision-making in multidisciplinary team-based cardiovascular care. JACC Adv. 2024;3:100981.
7. Bouniols N, Leclère B, Moret L. Evaluating the quality of shared decision making during the patient-carer encounter: a systematic review of tools. BMC Res Notes. 2016;9:382.
8. Gärtner FR, Bomhof-Roordink H, Smith IP, Scholl I, Stiggelbout AM, Pieterse AH. The quality of instruments to assess the process of shared decision making: a systematic review. PLoS ONE. 2018;13:e0191747.
9. Grant RW, Adams AS, Bayliss EA, Heisler M. Establishing visit priorities for complex patients: A summary of the literature and conceptual model to guide innovative interventions. Healthc (Amst). 2013;1:117–22.
10. Ng CJ, Lee PY. Practising shared decision making in primary care. Malays Fam Physician. 2021;16:2–7.
11. Braddock CH 3rd, Fihn SD, Levinson W, Jonsen AR, Pearlman RA. How doctors and patients discuss routine clinical decisions. informed decision making in the outpatient setting. J Gen Intern Med. 1997;12:339–45.
12. Braddock CH 3rd, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. JAMA. 1999;282:2313–20.
13. Butow P, Juraskova I, Chang S, Lopez A-L, Brown R, Bernhard J. Shared decision making coding systems: how do they compare in the oncology context? Patient Educ Couns. 2010;78:261–8.
14. Clayman ML, Makoul G, Harper MM, Koby DG, Williams AR. Development of a shared decision making coding system for analysis of patient-healthcare provider encounters. Patient Educ Couns. 2012;88:367–72.
15. Selvaraj SP, Yen RW, Forcino RC, Elwyn G. Automating the Observer OPTION-5 measure of shared decision making: assessing validity by comparing large language models to human ratings. Patient Educ Couns. 2025;142:109362.
16. Park YJ, Pillai A, Deng J, Guo E, Gupta M, Paget M, et al. Assessing the research landscape and clinical utility of large language models: a scoping review. BMC Med Inf Decis Mak. 2024;24:72.
17. Makoul G, Clayman ML. An integrative model of shared decision making in medical encounters. Patient Educ Couns. 2006;60:301–12.
18. U.S. Department of Veterans Affairs. Physician-patient transcripts with 4C coding analysis from the Contextualizing Care research program. VA open data portal. https://www.data.va.gov/dataset/Physician-patient-transcripts-with-4C-coding-analy/4qbs-wgct. Accessed 5 Apr 2026.

19. Weiner SJ, Kelly B, Ashley N, Binns-Calvey A, Sharma G, Schwartz A, et al. Content coding for contextualization of care: evaluating physician performance at patient-centered decision making. Med Decis Mak. 2014;34:97–106.
20. Guimond P, Bunn H, O'Connor AM, Jacobsen MJ, Tait VK, Drake ER, et al. Validation of a tool to assess health practitioners' decision support and communication skills. Patient Educ Couns. 2003;50:235–45.
21. Elwyn G, Edwards A, Wensing M, Hood K, Atwell C, Grol R. Shared decision making: developing the OPTION scale for measuring patient involvement. Qual Saf Health Care. 2003;12:93–9.
22. Peters M, Potter CM, Kelly L, Fitzpatrick R. Self-efficacy and health-related quality of life: a cross-sectional study of primary care patients with multi-morbidity. Health Qual Life Outcomes. 2019;17:37.
23. Wieringa TH, Rodriguez-Gutierrez R, Spencer-Bonilla G, de Wit M, Ponce OJ, Sanchez-Herrera MF, et al. Decision aids that facilitate elements of shared decision making in chronic illnesses: a systematic review. Syst Rev. 2019;8:121.
24. Chawla N, Arora NK. Why do some patients prefer to leave decisions up to the doctor: lack of self-efficacy or a matter of trust? J Cancer Surviv. 2013;7:592–601.
25. Elwyn G, Frosch DL, Kobrin S. Implementing shared decision-making: consider all the consequences. Implement Sci. 2016;11:114.

Figures

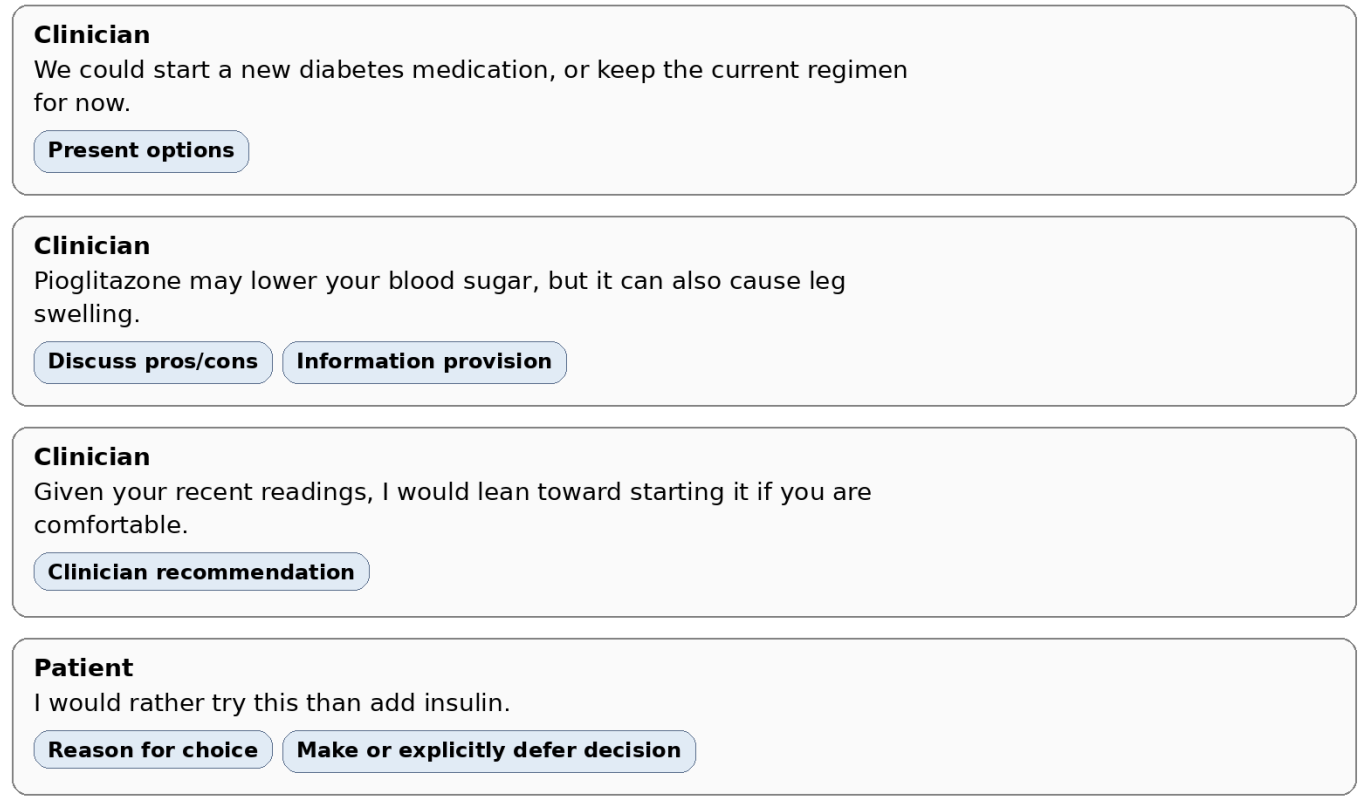


Figure 1

Illustrative example of utterance-level multi-label annotation of shared decision-making elements.

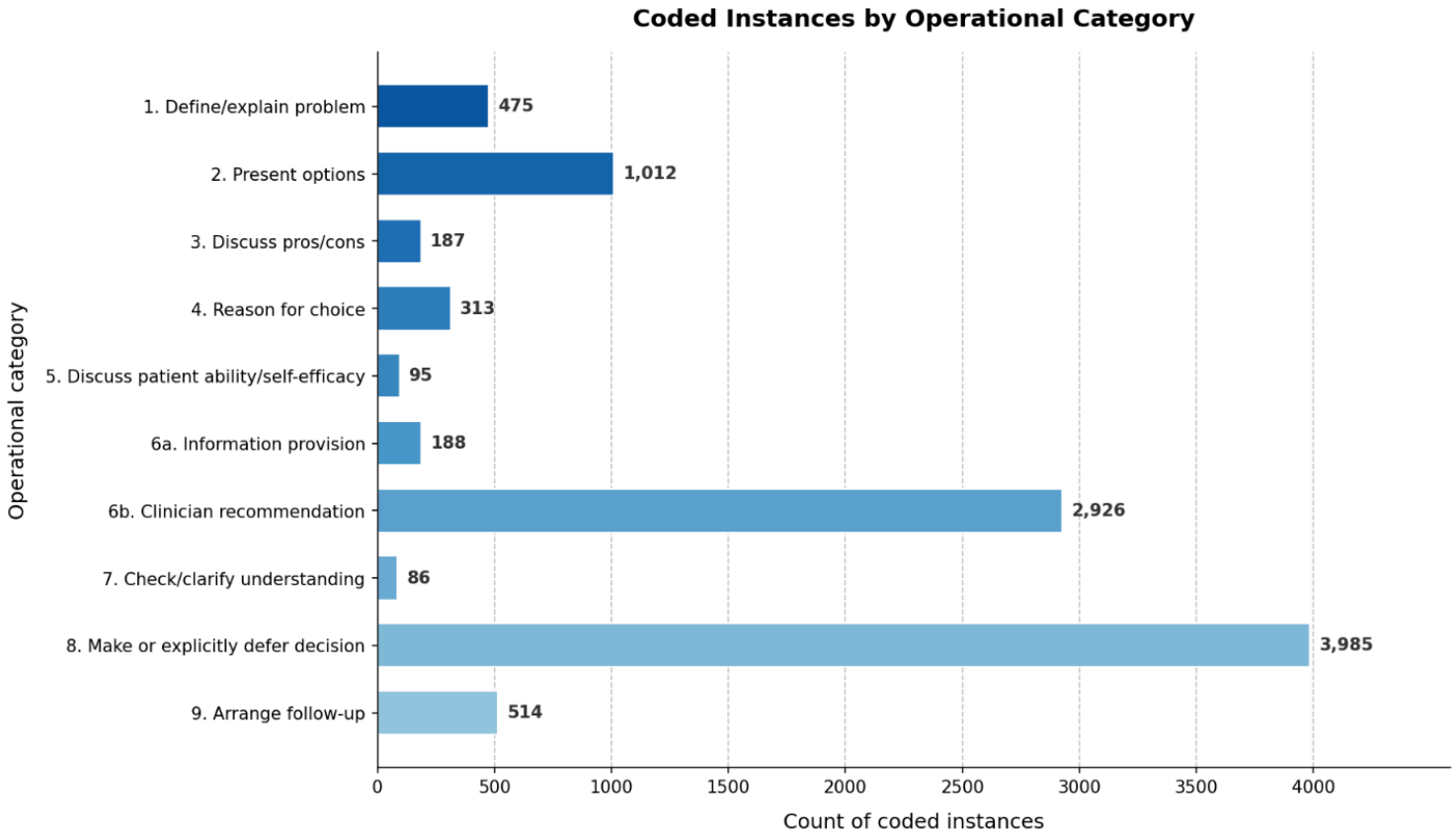


Figure 2

Frequency distribution of annotated shared decision-making codes across the corpus (coded instances, not visits).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx