UDC 811.111’42:61
DOI https://doi.org/10.32782/2710-4656/2026.2.1/22

Sivaieva O. S.
https://orcid.org/0000-0003-0123-9446
Borys Grinchenko Kyiv Metropolitan University

UNCERTAINTY IN MEDICAL TEXTS: A CORPUS-BASED ANALYSIS

The phenomenon of linguistic uncertainty is defined as a fundamental constituent of medical communication, necessitated by the inherently probabilistic nature of clinical evidence and therapeutic outcomes. This research provides an extensive, corpus-based investigation into the systematic encoding of such uncertainty across divergent medical genres. While the presence of hedging and modality in scientific prose has been previously studied, the specific sociolinguistic variations across professional, lay, and regulatory contexts remain under-documented. Consequently, the stylistic and pragmatic mechanisms of hedging are scrutinized to determine how medical knowledge is recontextualized for different stakeholders.

A specialized corpus of thirty English-language medical texts was compiled and categorized into three discrete genres: academic research articles (n=10), patient-oriented informational brochures (n=10), and pharmaceutical drug leaflets (n=10). The investigation was conducted via a dual-lens approach, where quantitative frequency distributions of epistemic markers were integrated with qualitative interpretations of their pragmatic functions. The interplay between modal verbs, lexical hedges, and adverbial qualifiers was examined to identify the transition from «cautious scientific detachment» to «directive risk communication».

The substantial variation in marker distribution is proven to be a direct function of the target audience and the specific communicative objective of the genre. These results are considered vital for the fields of medical translation, clinical communication, and health policy. By elucidating how medical «truth» is linguistically negotiated, this study contributes to a more robust framework for genre-sensitive linguistic mediation, ultimately enhancing the efficacy of health-related information dissemination.

Keywords: *uncertainty, medical discourse, corpus-based analysis, hedging, patient information texts, drug leaflets, risk communication.*

Statement of the problem. Medical communication operates in a domain inherently characterized by incomplete knowledge, evolving scientific evidence, and variable patient outcomes. As a result, uncertainty is not merely an occasional feature of medical texts but an integral structural property emerging from the nature of biomedical inquiry and clinical practice. From the cautious wording of academic research to the probabilistic risk categories in drug leaflets, linguistic strategies for expressing uncertainty play a crucial role in shaping how medical information is interpreted by experts, patients, and regulatory bodies.

Despite the centrality of uncertainty to medical discourse, its manifestation varies considerably across genres. Academic researchers use hedging to align claims with available evidence, mitigate potential criticism, and maintain epistemic humility. Patient-oriented texts attempt to simplify uncertainty without oversimplifying medical realities. Drug leaflets, in contrast, reflect legal and regulatory imperatives that require explicit descriptions of potential risks–even those unlikely to occur.

Analysis of recent research and publications. A foundational perspective is provided by Mishel, who describes uncertainty in health contexts as “the inability to determine the meaning of illness-related events”, emphasizing cognitive processes and the instability of available information [12, 227]. From a sociolinguistic standpoint, Zinn highlights the pervasive nature of uncertainty in modern risk societies, arguing that it is not merely a deficit but a structural condition of scientific and clinical practice [27, 441].

Within linguistic research, uncertainty is commonly divided into epistemic, aleatory, and pragmatic dimensions. Rowe and Wright define epistemic uncertainty as arising from incomplete or imperfect knowledge, noting that medical decisions often rely on probabilistic evidence rather than deterministic

truths [18, 344]. Aleatory uncertainty, in contrast, stems from inherent randomness and biological variability–what Gigerenzer et al. call "the natural limits of predictability" [5, 56]. Pragmatic uncertainty, discussed by Marková, refers to ambiguity or indeterminacy created through language itself, such as vague quantifiers or generalizing expressions [10, 150].

Several scholars emphasize that uncertainty is not solely a cognitive or informational problem but also a discursive phenomenon. For example, Han, Klein, and Arora propose a tripartite framework that distinguishes between sources, issues, and loci of uncertainty, illustrating how uncertainty is shaped by communication practices, institutional structures, and interpersonal dynamics [6, 831]. From the perspective of medical discourse analysis, Hyland demonstrates that uncertainty is rhetorically managed through hedging devices – lexical, grammatical, and pragmatic strategies that allow authors to present claims cautiously [7, 138]. According to Hyland, hedging in medical writing reflects disciplinary norms that prioritize precision and prevent overgeneralization.

Brashers argues that uncertainty can function as both a threat and a resource for patients: while it may increase anxiety, it can also preserve hope or enable flexibility in decision-making [2, 481]. This aligns with findings from patient communication studies, where Babrow shows that individuals interpret medical uncertainty through personal meaning-making frameworks rather than purely technical information [1, 557].

Finally, in the context of risk communication, Slovic and Peters et al. highlight how linguistic framing of uncertainty significantly shapes patient perception, often more than statistical evidence itself [20; 16]. Vague linguistic markers such as "may," "possible," "rare," or "in some cases" influence how individuals judge likelihood and severity of medical outcomes.

Task statement. This study aims to offer a systematic, corpus-based comparison of uncertainty marking across three genres of medical writing. Uncertainty in medical communication is a multidimensional construct that has been widely discussed in linguistics, sociology, psychology, and health communication studies. Its conceptualization varies across disciplines, but most scholars agree that uncertainty emerges from limitations in knowledge, inherent variability in biological systems, and communicative choices made by speakers.

Outline of the main material of the study. Taken together, the literature suggests that uncertainty in medical communication is a multilayered construct encompassing epistemic (knowledge-based), aleatory (variability-based), and pragmatic (language-based) dimensions. It is shaped not only by the state of scientific knowledge but also by institutional norms, communicative practices, and social expectations. This multidimensional perspective provides the conceptual foundation for exploring how uncertainty is linguistically encoded across different medical genres.

This study draws on a mini-corpus of thirty English medical texts published between 2018 and 2024. The corpus incorporates materials from three major genres – academic research articles, patient information texts, and pharmaceutical leaflets – to ensure balanced coverage across expert, semi-expert, and lay-oriented medical communication (See Table 1). All illustrative examples are paraphrased and represent recurrent linguistic patterns identified across the corpus rather than direct quotations from individual online sources.

In total, the corpus encompasses approximately 45,000 tokens. This size ensures adequate lexical density and statistical reliability for frequency-based measurements while remaining manageable for in-depth qualitative interpretation.

Three categories of uncertainty markers were examined:

1. Lexical markers – modal verbs, adverbs, adjectives, cognitive verbs.
2. Syntactic markers – conditional clauses, passive constructions, agentless passives.
3. Pragmatic markers – disclaimers, depersonalization, regulatory warnings.

The analysis combined corpus-driven quantitative methods with qualitative interpretation. Frequency counts and collocate analyses were generated using AntConc and subsequently refined through contextual analysis of individual occurrences. The table below reports normalized frequencies per 1,000 words.

Academic texts exhibit the highest density of epistemic uncertainty markers, highlighting the central role of hedging in the cautious presentation of scientific

Table 1

Overview of the Corpus Structure by Genre

Genre	Number of Texts	Approx. Word Count	Source Examples
Academic research articles	10	~ 22,000 words	PubMed, BMJ, JAMA
Patient information texts	10	~ 14,000 words	NHS, Mayo Clinic, WebMD
Drug leaflets	10	~ 9,000 words	FDA-approved inserts

Table 2

Distribution of Uncertainty Markers in the Medical Mini-Corpus

Marker Type	Academic Articles	Patient Texts	Drug Leaflets
Modal verbs	18.9	15.4	22.7
Epistemic adverbs	12.8	4.6	2.3
Cognitive verbs	10.3	3.4	1.7
Conditionals	7.1	6.2	13.4
Regulatory disclaimers	1.4	3.9	17.8

knowledge. Patient-oriented texts, by contrast, employ a smaller overall range of uncertainty markers, relying predominantly on modal verbs to express possibility and variability in an accessible manner for non-expert readers. Drug leaflets, however, are characterized by exceptionally high frequencies of conditional constructions and regulatory disclaimers, a pattern that directly reflects the legal and institutional requirements governing pharmaceutical communication.

In academic medical writing, uncertainty is primarily realized through epistemic hedging devices that allow authors to cautiously frame empirical claims. Previous corpus-based studies demonstrate that combinations of cognitive verbs and modal verbs are particularly frequent in this genre [7; 12]. Typical formulations include statements such as "the findings suggest that early intervention may reduce disease progression" or "it is plausible that genetic variability influenced the observed outcomes", patterns that have been widely documented in biomedical research articles [8; 14]. These constructions enable authors to calibrate the strength of their claims, explicitly acknowledge methodological limitations, and align their argumentation with disciplinary conventions that prioritize epistemic caution and scientific accountability [13, 198].

Patient-oriented medical texts employ a more restricted and simplified repertoire of uncertainty markers, relying predominantly on modal verbs to communicate potential risks and outcomes. Research on health communication consistently shows that expressions such as "the condition may cause mild discomfort in some individuals" or "symptoms can vary depending on overall health" are characteristic of patient information materials produced by institutions such as the NHS and the Mayo Clinic [13; 14; 9]. These formulations serve to communicate uncertainty transparently while avoiding deterministic or alarming language. From a pragmatic perspective, such markers support informed decision-making and align with principles of uncertainty management by reducing anxiety and maintaining patient trust [2; 21].

In pharmaceutical drug leaflets, uncertainty is expressed through highly standardized and legally mandated linguistic patterns. Regulatory guidelines issued by the European Medicines Agency and the U.S. Food and Drug Administration require the systematic use of modal verbs and probability labels in side-effect descriptions [3; 23]. Typical examples include formulations such as "this medication may cause uncommon but serious adverse reactions" or conditional warnings like "if swelling occurs, discontinue use and contact a physician", which have been identified as core features of patient information leaflets in multiple discourse-analytic studies [24; 4]. Additionally, side effects are routinely classified using standardized frequency categories (common, uncommon, rare, very rare), a practice designed to ensure legal compliance and uniform risk communication across pharmaceutical products [5, 91]. Functionally, these uncertainty markers fulfill regulatory obligations, standardize the presentation of medical risk, and shift responsibility toward patient compliance by explicitly linking uncertain outcomes to prescribed courses of action.

The quantitative results presented in Table 2 reveal systematic and statistically meaningful differences in the distribution of uncertainty markers across the three medical genres examined. These differences confirm that uncertainty marking is not a uniform linguistic phenomenon but a genre-sensitive communicative strategy shaped by epistemic goals, audience expectations, and institutional constraints.

The high frequencies of epistemic adverbs (12.8 per 1,000 words) and cognitive verbs (10.3 per 1,000 words) observed in academic research articles underscore the central role of epistemic hedging in scientific discourse. As illustrated by examples such as "the findings suggest that early intervention may reduce disease progression", academic authors consistently frame claims as provisional interpretations rather than definitive conclusions. This pattern aligns with established accounts of scientific writing as a socially regulated practice in which claims must be carefully calibrated to avoid overstatement [7; 12]. The results thus support the view that uncertainty in academic texts functions as a legitimizing mechanism, allowing researchers to negotiate knowledge claims while maintaining disciplinary credibility.

In contrast, patient-oriented texts exhibit a reduced overall density of uncertainty markers, particularly epistemic adverbs and cognitive verbs. Instead, modal verbs dominate this genre (15.4 per 1,000 words), as seen in formulations such as "the condition may cause mild discomfort" or "symptoms can vary depending on overall health." This distribution reflects a strategic simplification of uncertainty, whereby probabilistic information is conveyed through linguistically accessible forms. Rather than engaging in epistemic negotiation, patient texts prioritize clarity and emotional reassurance, supporting earlier findings that patient communication favors interpretability over analytical precision [15, 10]. The results therefore indicate that uncertainty in this genre serves a primarily interpersonal and pedagogical function, enabling informed yet non-alarming decision-making.

Drug leaflets demonstrate a markedly different uncertainty profile. The exceptionally high frequencies of conditional constructions (13.4 per 1,000 words) and regulatory disclaimers (17.8 per 1,000 words) highlight the institutionalized nature of uncertainty in pharmaceutical communication. Unlike academic and patient texts, where uncertainty reflects epistemic limitation or communicative choice, uncertainty in drug leaflets is imposed by regulatory frameworks. Modal verbs (22.7 per 1,000 words) further reinforce this pattern, functioning less as hedging devices and more as legally mandated indicators of potential risk. These findings support the argument that uncertainty in drug leaflets fulfills a protective legal function, ensuring compliance, standardization, and explicit risk disclosure.

Taken together, the results demonstrate that similar linguistic resources – such as modal verbs and conditionals – perform fundamentally different functions across genres. While they contribute to epistemic caution in academic writing, they facilitate accessibility in patient texts and enforce regulatory compliance in drug leaflets. The tight correspondence between quantitative frequencies and qualitative examples confirms the value of a corpus-based approach for uncovering genre-specific uncertainty strategies in medical discourse.

A cross-genre comparison of the findings demonstrates that uncertainty fulfills fundamentally different communicative functions across medical text types. In academic research articles, uncertainty primarily serves an epistemic function, enabling authors to express scientific caution and reinforce disciplinary credibility. Through hedging devices such as epistemic adverbs and cognitive verbs, researchers position their claims as provisional and open to further verification, thereby aligning with established norms of scientific accountability.

In patient-oriented texts, uncertainty performs a predominantly interpersonal and pedagogical role. Rather than supporting epistemic negotiation, uncertainty markers in this genre function to enhance clarity, provide reassurance, and ensure accessibility for non-expert readers. The reliance on modal verbs reflects an effort to communicate medical risk transparently while maintaining an empathetic tone that supports informed yet non-anxious decision-making.

By contrast, uncertainty in drug leaflets is shaped primarily by legal and regulatory imperatives. Here, uncertainty functions as a mechanism of risk notification and legal protection, ensuring compliance with institutional guidelines governing pharmaceutical communication. Conditional constructions, standardized probability labels, and regulatory disclaimers explicitly encode potential adverse outcomes and prescribe appropriate patient responses, thereby transferring responsibility for risk management to the user.

Taken together, this cross-genre comparison illustrates that uncertainty in medical communication is not merely a reflection of incomplete knowledge but a flexible communicative resource adapted to audience, purpose, and institutional context. Medical genres strategically recalibrate uncertainty to balance epistemic responsibility, communicative accessibility, and regulatory compliance, underscoring the need for genre-sensitive approaches in health communication research.

The analysis reveals that academic texts rely heavily on epistemic hedging devices to construct cautious, provisional claims consistent with scientific norms of evidence and accountability. Patient-oriented texts, while exhibiting fewer uncertainty markers overall, strategically employ accessible modal verbs to communicate risk and variability in a manner that supports comprehension and emotional reassurance. Drug leaflets, in contrast, encode uncertainty through standardized, regulatory-driven formulations that reflect legal obligations rather than epistemic choice.

Conclusions. These findings contribute to a growing body of research emphasizing the discursive nature of medical uncertainty. Rather than being a simple reflection of incomplete knowledge, uncertainty emerges as a communicative resource shaped by genre, audience, and institutional context. From an applied perspective, the results have implications for health communication, medical translation, and the design of patient information materials, suggesting that effective uncertainty management requires genre-sensitive linguistic strategies.

Future research could extend this analysis by incorporating larger corpora, cross-linguistic comparisons, or multimodal elements such as visual risk representations. Nevertheless, the present study demonstrates that a focused corpus-based approach offers valuable insights into how uncertainty is constructed, negotiated, and institutionalized within contemporary medical communication.

Bibliography:

1. Babrow A. S. Uncertainty, value, communication, and problematic integration. *Journal of Communication*. 2001. Vol. 51, № 3. P. 553–573.
2. Brashers D. E. Communication and uncertainty management. *Journal of Communication*. 2001. Vol. 51, № 3. P. 477–497.
3. European Medicines Agency. Guideline on the readability of the labelling and package leaflet of medicinal products for human use. London : EMA, 2009.
4. Fage-Butler A. M., Jensen M. N. The rhetorical construction of uncertainty in patient information leaflets. *Health Communication*. 2015. Vol. 30, № 4. P. 372–383.
5. Gigerenzer G., Gaissmaier W., Kurz-Milcke E., Schwartz L. M., Woloshin S. Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest.* 2007. Vol. 8, № 2. P. 53–96.
6. Han P. K. J., Klein W. M. P., Arora N. K. Varieties of uncertainty in health care: A conceptual taxonomy. *Medical Decision Making*. 2011. Vol. 31, № 6. P. 828–838.
7. Hyland K. Hedging in scientific research articles. – Amsterdam : John Benjamins Publishing Company, 1998. – 272 p.
8. Hyland K. Metadiscourse: Exploring interaction in writing. – London : Continuum, 2005. – 230 p.
9. Kandula N. R., Bennett C. L., Nápoles A. M. Literacy and patient understanding of medical information. *Journal of General Internal Medicine*. 2010. Vol. 25, № 2. P. 169–174.
10. Marková I. The dialogical mind: Common sense and ethics. Cambridge : Cambridge University Press, 2017. 284 p.
11. Mayo Clinic : patient education resources. URL: https://www.mayoclinic.org
12. Mishel M. H. Uncertainty in illness. Image: The Journal of Nursing Scholarship. 1988. Vol. 20, № 4. P. 225–232.
13. Myers G. Writing biology: Texts in the social construction of scientific knowledge. Madison : University of Wisconsin Press, 1990. – 312 p.
14. National Health Service (NHS) : patient information portal. URL: Mode of access: https://www.nhs.uk
15. Pander Maat H., Lentz L. Improving clarity of patient information leaflets. *Journal of Health Communication*. 2010. Vol. 15, № 1. P. 1–20.
16. Peters E., Hibbard J., Slovic P., Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs*. 2006. Vol. 25, № 3. P. 741–748.
17. PubMed : database of biomedical literature URL: https://pubmed.ncbi.nlm.nih.gov
18. Rowe G., Wright G. Differences in expert and lay judgments of risk. *Risk Analysis*. 2001. Vol. 21, № 2. P. 341–356.
19. Salager-Meyer F. Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*. 1994. Vol. 13, № 2. P. 149–170.
20. Slovic P. The perception of risk. London : Earthscan, 2000. 473 p.
21. Thompson T., Hunt J. Managing uncertainty in patient-centered communication. *Health Communication*. 2018. Vol. 33, № 3. P. 287–295.
22. U.S. Food and Drug Administration (FDA) : drug labels and package inserts URL: Mode of access: https://www.accessdata.fda.gov
23. U.S. Food and Drug Administration. Drug labeling requirements. – Silver Spring : FDA, 2020.
24. Van Bree J., Pander Maat H. Risk communication in pharmaceutical package leaflets. *Technical Communication*. 2011. Vol. 58, № 3. P. 229–245.
25. Varttala T. Hedging in scientific research articles: A cross-disciplinary study. *English for Specific Purposes*. 2001. Vol. 20, № 4. P. 383–400.
26. WebMD : health information for consumers URL: https://www.webmd.com
27. Zinn J. O. Heading into the unknown: Everyday strategies for managing risk and uncertainty. *Health, Risk & Society*. 2008. Vol. 10, № 5. P. 439–450.

Сіваєва О. С. НЕВПЕВНЕНІСТЬ У МЕДИЧНОМУ ДИСКУРСІ: КОРПУСНО ОРІЄНТОВАНИЙ АНАЛІЗ

Феномен лінгвістичної невизначеності визначається як фундаментальний складник медичної комунікації, зумовлений іманентно пробабілістичною природою клінічних доказів та терапевтичних результатів. У представленому дослідженні здійснено комплексний корпусно-орієнтований аналіз системного кодування такої невизначеності в межах різних медичних жанрів. Попри те, що прояви хеджування та модальності в наукових текстах уже були вивчені раніше, специфічні соціолінгвістичні варіації в професійному, аматорському та регуляторному контекстах залишаються недостатньо задокументованими. Відповідно, стилістичні та прагматичні механізми хеджування піддано критичному аналізу з метою з'ясування особливостей реконтекстуалізації медичних знань для різних стейкголдерів.

Для проведення дослідження сформовано спеціалізований корпус із тридцяти англомовних медичних текстів, які були класифіковані за трьома дискретними жанрами: академічні наукові статті (n=10), інформаційні брошури для пацієнтів (n=10) та інструкції до лікарських засобів (n=10). Дослідження проведено із застосуванням двоаспектного підходу, у межах якого кількісні показники розподілу епістемічних маркерів інтегровано з якісною інтерпретацією їхніх прагматичних функцій. Взаємодію модальних дієслів, лексичних хеджів та прислівникових кваліфікаторів досліджено для ідентифікації переходу від стратегії «обережної наукової відстороненості» до «директивної комунікації ризиків».

Суттєву варіативність у дистрибуції маркерів доведено як пряму функцію цільової аудиторії та конкретної комунікативної мети жанру. Отримані результати вважаються визначальними для галузей медичного перекладу, клінічної комунікації та політики охорони здоров'я. Шляхом з'ясування механізмів лінгвістичного узгодження медичної «істини» дане дослідження робить внесок у розробку обґрунтованої моделі жанрово-чутливої мовної медіації, що в перспективі сприятиме підвищенню ефективності поширення медичної інформації.

Ключові слова: *невпевненість, медичний дискурс, корпусний аналіз, хеджування, інформаційні тексти для пацієнтів, листки-вкладиші до препаратів, комунікація ризиків.*

Дата першого надходження статті до видання: 23.02.2026
Дата прийняття статті до друку після рецензування: 20.03.2026
Дата публікації (оприлюднення) статті: 14.05.2026