oncOPAL — Oncology Optimized Patient-Accessible Language

Study Protocol

Prospective Randomized Controlled Trial to Evaluate Locally Implemented Large Language Models (LLMs) for Simplifying Patient Communication in Hematology and Oncology

Date: 18.03.2026

Version: 1.0

Study Code: oncOPAL

Principal Investigators

Dr. med. Krischan Braitsch

PD Dr. med. Lisa C. Adams

TUM University Hospital (Klinikum rechts der Isar)

Ismaninger Str. 22, 81675 Munich

1 Protocol Synopsis

Short Title	oncOPAL
Title	Prospective Randomized Controlled Trial to Evaluate Locally Implemented Large Language Models (LLMs) for Simplifying Patient Communication in Hematology and Oncology
Study Center	TUM University Hospital (Klinikum rechts der Isar)
Study Design	Prospective, randomized controlled trial with parallel group design. Block randomization 2:1 (intervention : control) with permuted blocks of variable size (4–6). Additionally, a non-randomized translation arm for non-German-speaking patients.
Study Population	Inpatients of the Department of Medicine III (Hematology/Oncology) at TUM University Hospital who receive a discharge letter including the sections Current Status, Medical History, Epicrisis, and Further Management as part of routine clinical care.
Rationale	Up to 40–80% of medical information is not correctly recalled or understood by patients. LLMs can convert medical texts into plain-language versions with high accuracy. However, prospective RCTs evaluating the clinical benefit of LLM-simplified patient synopses are lacking.
Primary Objective	To evaluate the superiority of LLM-generated, simplified patient synopses over standard synopses with respect to patient comprehension.
Intervention	Intervention group: The following sections of the discharge letter are simplified by the LLM system: Current Status, Medical History, Epicrisis, and Further Management. The patient receives the LLM-

	simplified version in addition to the standard synopsis (after physician review). Control group: Standard synopsis only. Translation arm: LLM-simplified and translated synopsis of the same sections.
Sample Size	180 patients (150 randomized: 100 intervention, 50 control; 30 translation arm)
Primary Endpoint	Comprehension score (5-item scale, 10-point Likert, based on PEMAT)
Secondary Endpoints	Patient satisfaction (EORTC QLQ-INFO25 subscales), uncertainty reduction, format preference, physician review workload, correction rate, translation quality
Statistical Analyses	t-test / Mann-Whitney U-test for primary endpoint; chi-squared / Fisher for categorical endpoints; multivariate regression adjusted for educational level and further covariates; ITT analysis. Analysis with R or SPSS.
Study Duration	12 months recruitment (April 2026 – March 2027), data analysis complete June 2027

2 Rationale

2.1 Background

Communicating complex medical information to patients represents a central challenge in oncology. Studies show that up to 40–80% of the medical information conveyed during a physician consultation is not correctly recalled or understood by patients (Kessels 2003). This is particularly relevant in hematology and oncology, where complex treatment regimens, prognoses, and side-effect profiles must be communicated, making comprehensible written summaries especially important.

Large language models (LLMs) have made substantial progress in processing and generating natural language over recent years. Current studies demonstrate that LLMs can translate medical technical texts into plain language with high accuracy (Jeblick et al. 2024). A systematic review by Ayers et al. (2023) showed that LLM-generated responses to patient questions were in some cases rated higher in quality and empathy than responses from physicians.

Local implementation of LLMs (on-premise) enables the processing of sensitive patient data in compliance with the GDPR, as no data are transmitted to external servers.

2.2 Derivation of the Research Question

Despite promising results in retrospective analyses, prospective, randomized controlled trials evaluating the benefit of LLM-simplified patient synopses in routine clinical practice are lacking. The present study addresses this research gap. In addition, a non-randomized arm will evaluate the feasibility and quality of LLM translations for patients with insufficient German language proficiency.

3 Study Objectives

3.1 Primary Objective

To evaluate the superiority of LLM-generated, simplified patient synopses over standard synopses with respect to patient comprehension in hematology and oncology.

3.2 Secondary Objectives

- Assessment of patient satisfaction with the information received
- Evaluation of the subjective reduction in uncertainty and anxiety
- Assessment of patient preference for synopsis format
- Measurement of the time required for physician review of LLM-generated synopses
- Documentation of the rate of necessary physician corrections
- Qualitative assessment of LLM translations by native-speaker reviewers

3.3 Endpoints

Primary Endpoint:

Comprehension score of the patient synopsis, measured using a 5-item scale (10-point Likert, based on PEMAT). The score is calculated as the mean of five items: overall comprehension, comprehension of the diagnosis, the treatment, the next steps, and medical terminology.

Secondary Endpoints:

- Patient satisfaction (EORTC QLQ-INFO25 subscales)
- Subjective uncertainty reduction (single item, 0–10 scale, measured before and after reading the synopsis)
- Preference for synopsis format (categorical variable)
- Health literacy as a covariate (HLS-EU-Q6)
- Time for physician review (in minutes) and correction rate
- Translation quality (rated by native-speaker reviewers, 5-point scale)

3.4 Hypotheses

H1: LLM-simplified patient synopses achieve a significantly higher comprehension score (difference of at least 1.5 points on the 10-point scale) compared to standard synopses.

H0: There is no difference or a difference of less than 1.5 points between the groups.

4 Study Procedures

4.1 Study Design

Prospective, single-center, randomized controlled trial with parallel group design. Randomization is performed in a 2:1 ratio (intervention : control) using a computer-generated randomization list with permuted blocks of variable size (4–6). An additional non-randomized translation arm is conducted for non-German-speaking patients.

Blinding is not possible, as the intervention group receives an additional, visibly different synopsis. However, questionnaire evaluation is performed blinded to group assignment.

4.2 Study Procedures

Potential participants are identified during inpatient stays in the Department of Medicine III. Following verification of inclusion and exclusion criteria, informed consent is obtained in writing (with a reflection period of at least 24 hours). Randomization or assignment to the translation arm then follows.

Day 0 (as part of the regular hospital visit):

- Informed consent procedure (approx. 10–15 minutes)
- Randomization or assignment to the translation arm
- Preparation of the synopsis as part of routine clinical care
- LLM simplification and physician review where applicable
- Distribution of synopsis/synopses with at least 15 minutes reading time
- Completion of the questionnaire (approx. 5–10 minutes)

No further study visits are planned. Study participation is complete upon submission of the questionnaire.

4.3 Intervention

Intervention Group:

The following sections of the discharge letter are automatically converted into a plain-language version by the locally implemented LLM system (GPT-OSS, on-premise): Current Status, Medical History, Epicrisis, and Further Management. A study physician reviews the LLM-generated synopsis for content accuracy and makes corrections as necessary. The patient receives the simplified version in addition to the standard synopsis.

Control Group:

The patient receives the standard synopsis only.

Translation Arm:

The specified sections of the discharge letter (Current Status, Medical History, Epicrisis, Further Management) are simplified by the LLM system and translated into the patient's native language. The translation is reviewed by a native-speaker reviewer.

4.4 Technical Implementation

The LLM system (GPT-OSS) is operated locally on servers at the Klinikum rechts der Isar (on-premise). No data are transmitted to external servers or cloud services. The system is not classified as a medical device and is used solely to support text simplification, not for diagnosis or treatment decisions.

5 Criteria for Early Study Termination

Individual Discontinuation:

- Withdrawal of consent by the participant
- Subsequent identification of an exclusion criterion

Termination of the Entire Study:

- Decision by the principal investigators for scientific or organizational reasons
- Request by the responsible ethics committee or supervisory authority
- Unexpected, serious safety concerns

6 Study Population

6.1 Inclusion Criteria

- Age 18 years or older
- Inpatients of the Department of Medicine III (Hematology/Oncology) at TUM University Hospital
- Receipt of a discharge letter including the sections Current Status, Medical History, Epicrisis, and Further Management as part of routine clinical care
- Capacity to provide informed consent
- Written informed consent following the consent procedure

Additional Inclusion Criteria for the Translation Arm:

- Insufficient German language proficiency
- Native language supported by the LLM system (English, Turkish, Arabic, Russian, Polish, Spanish, Italian, French, Portuguese, Vietnamese, Chinese)

6.2 Exclusion Criteria

- Cognitive impairments that preclude independent assessment of comprehension (e.g., dementia, severe encephalopathy)
- Participation in another study with potential influence on the endpoints
- Lack of capacity to provide informed consent
- Refusal to participate in the study

6.3 Patient Recruitment

Recruitment proceeds consecutively during inpatient ward rounds in the Department of Medicine III. Potential participants are identified by the treating medical staff. No recruitment materials such as flyers or notices are used.

7 Data Protection and Data Security

7.1 Informed Consent

Study participation requires prior written informed consent. The consent procedure is conducted by a study physician. Participants are given a reflection period of at least 24 hours. Consent may be withdrawn at any time without giving reasons and without any disadvantage to further treatment.

7.2 Legal Basis

The legal basis for processing personal data is informed consent pursuant to Art. 6(1)(a) and Art. 9(2)(a) GDPR in conjunction with § 27 BDSG.

7.3 Data Collection and Storage

Data collected include: demographic data (age, sex, educational level, native language), clinical baseline data (diagnosis, treatment setting), questionnaire data, and interview data where applicable. All study data are pseudonymized (study code oncOPAL-XXXX) and stored in a password-protected electronic database on servers of the Klinikum rechts der Isar. Data transmission is encrypted (TLS 1.3).

7.4 Pseudonymization

Each participant is assigned a unique alphanumeric study code. The mapping list is stored separately from the study data in an encrypted document. Access is restricted to the principal investigator and a designated deputy. All analyses are performed exclusively with pseudonymized data.

7.5 Data Disclosure

Personal data are not disclosed to third parties. Publications use exclusively aggregated, anonymized data. Transfer of data to non-EU countries is not planned.

7.6 Withdrawal and Data Deletion

In the event of withdrawal, no further data are collected. Already collected data may be deleted upon request, provided they have not yet been incorporated into the analysis in anonymized form. After expiry of the retention period (10 years), all study data will be fully deleted and the mapping list destroyed.

7.7 Access Control

Access to study data is protected by individual user accounts with strong passwords and is logged. Servers are located in a secured data center with access control.

8 Benefit-Risk Assessment

The benefit-risk ratio is favorable. Participants in the intervention group and translation arm receive an additional plain-language synopsis that can promote understanding of diagnosis and treatment. Study participation involves minimal burden (additional time requirement of approximately 5–15 minutes). There are no physical risks. Psychological risks are low; the risk of possible misunderstandings arising from the simplified presentation is mitigated by mandatory physician review. Routine clinical care remains completely unchanged.

9 Sample Size and Power Analysis

The sample size calculation is based on the following assumptions: a clinically relevant difference in the primary endpoint of 1.5 points, an expected standard deviation of 2.5 points, power of 80%, significance level $\alpha = 0.05$ (two-sided), a randomization ratio of 2:1, and an expected dropout rate of 10%.

For a two-sided t-test for independent samples, the required sample size is n = 136 (91 intervention, 45 control). Accounting for the dropout rate, n = 150 patients will be recruited for the randomized part.

For the translation arm, n = 30 patients will be enrolled. This sample size permits a descriptive evaluation of feasibility without a formal statistical comparison.

10 Funding

The study is funded through internal resources of the Institute for Diagnostic and Interventional Radiology and the Department of Medicine III at TUM University Hospital (Klinikum rechts der Isar). There are no conflicts of interest.

11 References

1. Kessels RPC. Patients' memory for medical information. J R Soc Med. 2003;96(5):219-222.

2. Jeblick K, et al. ChatGPT Makes Medicine Easy to Swallow: An Exploratory Case Study on Simplified Radiology Reports. Radiology. 2024;308(1):e230618.

3. Ayers JW, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Intern Med. 2023;183(6):589-596.

4. Arraras JI, et al. The EORTC QLQ-INFO25: A questionnaire to assess information given to cancer patients. Eur J Cancer. 2010;46(15):2726-2738.

5. Basch E, et al. Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. JAMA. 2017;318(2):197-198.