




Assessing the Accuracy of Symptoms and Adverse Events Reporting for Lung Cancer Treatment in the Danish National Patient Registry

Karoline Lindgaard Dybdahl ¹, Julie Krøier Bak¹, Emma Ekroll², Matilde Koldsø Søreid², Daniel Faurholdt Hansen²⁻⁴, Weronika Maria Szejniuk^{2,4,5}, Martin Bøgsted ²⁻⁴, Charles Vesteghem ²⁻⁴

¹Department of Health Science and Technology, Aalborg University, Gistrup, Denmark; ²Department of Clinical Medicine, Aalborg University, Aalborg, Denmark; ³Center for Clinical Data Science, Aalborg University and Aalborg University Hospital, Aalborg, Denmark; ⁴Clinical Cancer Research Centre, Aalborg University Hospital, Aalborg, Denmark; ⁵Department of Oncology, Aalborg University Hospital, Aalborg, Denmark

Correspondence: Karoline Lindgaard Dybdahl, Department of Health Science and Technology, Aalborg University, Selma Lagerlöfs Vej 249, Gistrup, 9260, Denmark, Email kldy21@student.aau.dk

Background: Systemic anti-cancer and radiation therapies for lung cancer often cause adverse events that impair quality of life. Therefore, it may be counterproductive if patient survival is too short for any benefit. Decision support tools predicting short-term mortality in patients with lung cancer have been investigated using data from the Danish National Patient Registry (DNPR). However, the extent of the DNPR's information on clinical manifestations of anti-cancer treatments, particularly in lung cancer, remains unknown. We assessed the validity of clinical manifestation reporting in the DNPR to evaluate its usability in decision support tools and epidemiological studies.

Patients and Methods: Ninety-five patients treated for lung cancer at the Department of Oncology, Aalborg University Hospital, who died between 2015 and 2022 were randomly selected. Four raters independently extracted clinical manifestations from patients' journals. Interrater agreement was measured using Light's kappa (threshold ≥ 0.41). Extracted manifestations were matched to DNPR data, and concordance was quantified using the F1-score, aggregating by date, week, month, quarter, year, and patient.

Results: Interrater agreement ranged from moderate to perfect, except for constipation. Concordance between journal and DNPR data was low for most clinical manifestations. F1-scores for dyspnea, hemoptysis, and pain, almost exclusively at the patient level, suggested moderate reporting in the DNPR.

Conclusion: Concordance between the DNPR and patient journals was limited for minor and moderate clinical manifestations, but severe manifestations, such as dyspnea, hemoptysis, and pain, showed moderate agreement, showcasing some usability in this limited context. This study also demonstrates the availability of rich data about clinical manifestations in journal notes. Despite potential biases in these data, this study highlights the potential in the extraction of clinical manifestations directly from patient journals, notably using Natural Language Processing, to obtain more detailed data for epidemiological studies or to build machine learning-based decision support tools.

Keywords: Danish national patient registry, systemic anti-cancer therapies, oncology, clinical manifestations, information extraction, journal notes

Background

Lung cancer is one of the most frequently diagnosed cancers worldwide.¹ In Denmark between 2017 and 2022, an average of 71.3 per 100,000 people were each year diagnosed with lung cancer and an average of 49.9 per 100,000 deaths per year due to lung cancer were reported.²

Currently available treatments for lung cancer include systemic anti-cancer therapies (SACTs), surgery, radiation therapy (RT), or a combination of these.^{3,4} In Denmark, the treatment strategy depends on the type of lung cancer, stage,

and patient performance status.⁵ Adverse events of SACTs depend on the treatment modality, whereas those of RT depend on the treatment site. Commonly observed adverse events of SACTs include fatigue, nausea, vomiting, hair loss, fever, and nerve damage, whereas the most common adverse events of RT are fatigue and symptoms of tissue irritation from the irradiated area.^{6,7} These adverse events are detrimental to the quality of life of patients and can even be counterproductive if the patient does not survive long enough to benefit from the treatment.

To assist clinicians in the decision-making processes, the development of decision support tools for predicting the short-term mortality of patients with lung cancer has been investigated.⁸ This decision support tool leverages extensive medical data to provide data-driven predictions including symptoms and adverse events. Data on symptoms and adverse events were obtained from the Danish National Patient Registry (DNPR).⁹ DNPR contains hospital administrative data, including diagnoses and treatments, and is widely used in epidemiological studies. Its validity has been demonstrated for SACTs,¹⁰ but it is unknown how much information concerning the symptoms and adverse events of anti-cancer treatments is available, notably for lung cancer.

In this study, we assessed the validity of reporting symptoms and adverse events in the DNPR to evaluate the usability of these data in decision support tools but also more broadly for epidemiological studies.

Methods and Materials

Patient Selection

The initial patient cohort consisted of 2380 lung cancer patients in contact with the Department of Oncology, Aalborg University Hospital, between 1991 and 2022, and with a documented date of death between 2015 and 2022. This ensures at least 24 years of potential history for each patient, hereby minimizing potential survival bias. Of the 2380 patients, 450 were excluded because they did not receive any form of anti-cancer treatment, defined as chemotherapy, radiotherapy, or surgery, leaving 1930 patients. These 1930 patients are referred to as the original cohort in the following sections.

Ninety-nine patients were randomly selected from the original cohort for further analyses of the patient journals.

Upon further inspection, four patients were excluded due to anti-cancer treatment for different cancer types. Consequently, this study was based on patient journals of 95 patients, referred to as the study population.

Pseudonymization was performed by replacing each patient's civil registration number (CPR number) with an incremental patient ID. In addition, names were removed from the text.⁹

Data Extraction

Patient Journals

Data on lung cancer patients were extracted from pseudonymized patient journals by four raters. The raters were undergraduate students in Medicine and Medicine with Industrial Specialization at Aalborg University. The rating was performed under the supervision of an experienced oncologist at Aalborg University Hospital. The extracted data, referred to as journal data, consisted of patient ID, predefined category, the exact wording of how the category was described, grading, and date of occurrence. The exact wording was used to assign the entry to one of the eleven categories of symptoms and adverse events, hereafter referred to as clinical manifestations, which were selected based on their relevance and frequency in a randomly selected subset of 12 patient journals from the study population.

Twelve journals were selected to ensure equal distribution among the four raters and to maintain a sample size above ten journals, which was considered sufficient to identify common clinical manifestations in the patient journals. The 11 clinical manifestations included constipation, coughing, diarrhea, dyspnea, fatigue, hemoptysis, headache, hoarseness, nausea, pain, and weight loss.

The grading was performed according to international guidelines of the Common Terminology Criteria for Adverse Events (CTCAE)¹¹ with the purpose of standardizing the collection of data. The grading guidelines are listed in [Supplementary Table 1](#). The data from 95 patients in the study population were extracted using this approach. Four patients from the study population were randomly chosen for review of their journal notes by all four raters. The remaining 91 journals were split as equally as possible among raters.

Patient Administrative System

The DNPR contains patient information notably diagnosis codes following the Danish Medical Classification System (SKS).¹² The codes used to identify clinical manifestations, as defined above, are listed in [Supplementary Table 2](#). Data to populate the DNPR are provided by each Danish region. In the North Denmark Region (NDR), these data are from the Patient Administrative System (PAS) which was the data source used in this study.

As no grading was available from the DNPR data, only patient ID, clinical manifestations, and date were extracted. Thus, the registered symptoms and side effects were graded as > 0 and considered positive compared with the grading system of the extracted data.

Analysis

Interrater Agreement Analysis

Interrater agreement was evaluated between the four raters using Light's kappa.¹³ To evaluate this, data entries from two raters at a time were matched based on the patient ID, clinical manifestation, and date. In cases where multiple assessments by the same rater for the same clinical manifestation were performed on the same day, only the assessment with the highest grading was retained. If a rater did not report any assessment on a day for a clinical manifestation, while other raters did, a value of -1 was used as grading for this rater. Confidence intervals for Light's kappa were calculated by bootstrapping by resampling with replacement of the identified occurrences and their corresponding six possible dependent pairwise assessments 1000 times. Light's kappa is an extension of Cohen's kappa for scenarios involving more than two raters. Therefore, a standardized interpretation of kappa values was used to evaluate the results.¹⁴ Light's kappa was interpreted as follows: Light's kappa < 0.00 indicates poor agreement, $0.00-0.20$ slight agreement, $0.21-0.40$ fair agreement, $0.41-0.60$ moderate agreement, $0.61-0.80$ substantial agreement, $0.81-1.00$ almost perfect agreement, and 1.00 perfect agreement.^{14,15} This interpretation allows for Light's kappa > 0.41 to be considered acceptable.

Danish National Patient Registry Compared to Patient Journals

For entries manually extracted from the four patients shared across raters, a consensus among raters was used. This consensus was obtained by comparing journal ratings and grading of clinical manifestations based on the majority of equal gradings. The journals reviewed by only one rater were deemed accurate upon achieving an acceptable level of interrater agreement.

To evaluate the concordance between the DNPR and journal data, records from both datasets were generated, consisting of patient ID, clinical manifestation, and date. For each patient and clinical manifestation, at most one record per date was kept in both datasets, hereafter referred to as an occurrence.

Six time period aggregations were considered: aggregating on date (no aggregation), week, month, quarter, year, and patient (no temporal criteria) and keeping only one record per aggregated time period. Both datasets were subsequently matched for patient ID, clinical manifestations, and aggregated time periods. For example, if a patient had a DNPR entry for fatigue on Monday the 23rd of March 2015, and fatigue was described in the patient journal on Tuesday the 31st of March 2015, this would count as a month, quarter, year, and patient match, but not a date and week match.

A matched DNPR record was considered a true positive, while a DNPR record that was not matched in the journal data was considered a false positive, and a non-matched journal data record was considered a false negative. In this setup, true negatives cannot be assessed.

Concordance was measured using the F1-score, which represents the harmonic mean between the sensitivity and the positive predictive value.¹⁶

$$F_1 score = \frac{2}{\frac{1}{PPV} + \frac{1}{Sensitivity}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Python 3.11 in Jupyter Notebook was used for data setup and analysis.

Approval

The Regional Council of the NDR approved this study under the Danish Health Act § 46, sec 2, under reference number 2022-044174. The study was registered in the NDR's research project inventory as F2022-186 to ensure compliance with Article 30 of the Danish Data Protection Regulation.

Results

Study Population

The characteristics of the study population are summarized in Table 1. The original cohort and study population exhibited similar distribution. Both were slightly overrepresented by men, with most patients above 60 years of age at the time of diagnosis (87% of patients). The majority of the patients in all considered cohorts survived at least 1 year from first diagnosis. Regarding treatment, SACTs alone or in combination with other therapies were the most commonly observed (69% of patients).

Table 1 Study Population Characteristics

	Original Cohort n, (%)	Exploratory Population n, (%)	Study Population n, (%)
Patients	1,930 (100)	12 (100)	95 (100)
Sex			
Male	997 (52)	6 (50)	52 (55)
Female	933 (48)	6 (50)	43 (45)
Diagnosis age, years			
18–59	256 (13)	≤5 (≤42)	17 (18)
60–74	1,033 (54)	8 (67)	45 (47)
≥75	641 (33)	≤5 (≤42)	33 (35)
Survival			
≥1 year	995 (52)	7 (58)	50 (53)
≥2 years	555 (29)	≤5 (≤42)	28 (29)
≥5 years	168 (9)	≤5 (≤42)	10 (11)
Anti-cancer treatment			
SACTs, RT, and surgery	367 (19)	≤5 (≤42)	24 (25)
SACTs and RT	479 (25)	0 (0)	21 (22)
SACTs and surgery	173 (9)	≤5 (≤42)	10 (11)
SACTs only	301 (16)	0 (0)	20 (21)
RT and surgery	140 (7)	≤5 (≤42)	≤5 (≤5)
RT only	262 (14)	0 (0)	11 (12)
Surgery only	208 (11)	≤5 (≤42)	≤5 (≤5)

Notes: Number of patients in the original cohort, exploratory population, and study population. The values in parenthesis represent the percentage of patients by sex, age at diagnosis interval, and type(s) of anti-cancer treatment(s), relative to the population. The survival is calculated from first diagnosis.

Abbreviations: n, number of patients; SACTs, systemic anti-cancer therapies; RT, radiation therapy.

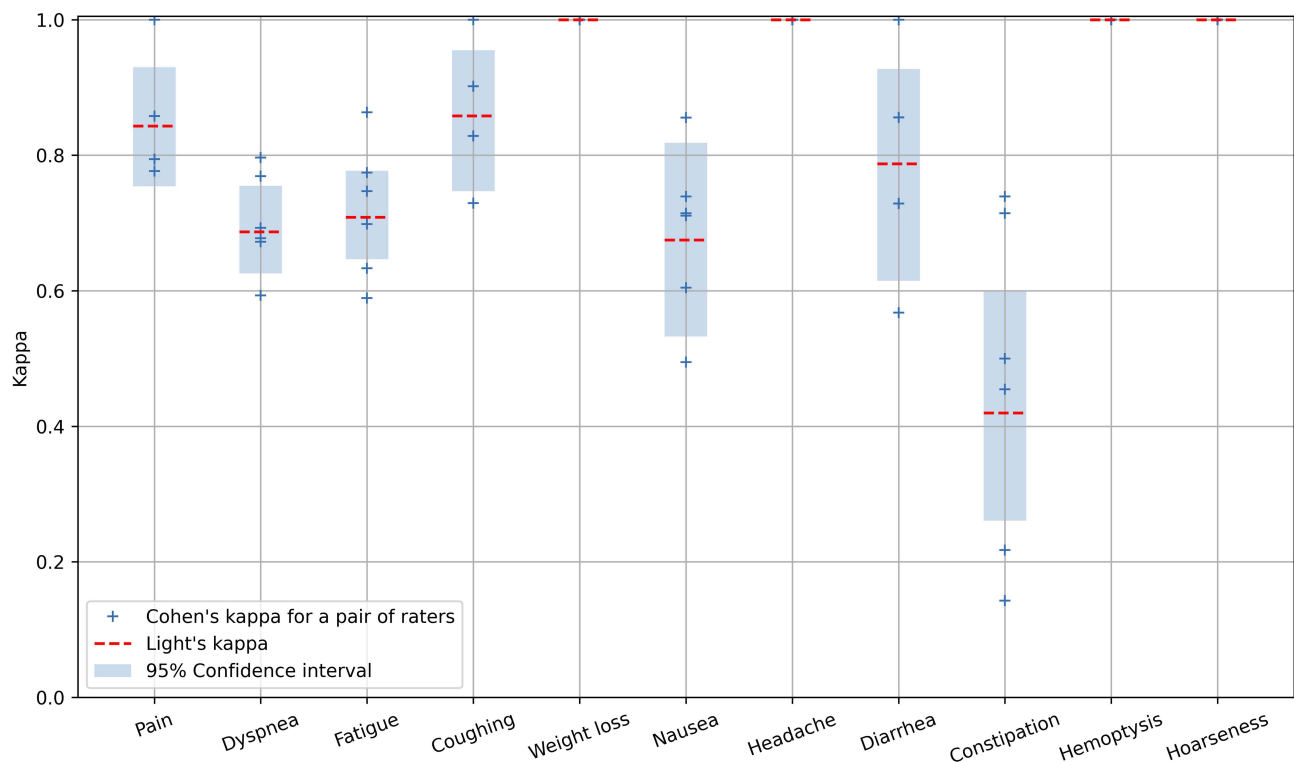


Figure 1 Interrater agreement analysis displayed by Light's Kappa between all four raters for each clinical manifestation with 95% confidence intervals. Confidence intervals were obtained by bootstrapping and are displayed as blue bars around the red striped lines showing the mean Light's kappa.

Interrater Agreement Analysis

Figure 1 shows the mean results of the interrater agreement analysis. The analysis compared the four raters' assessments of clinical manifestations. The results showed that Light's kappa for all clinical manifestations exceeded 0.65, except for constipation, which had a Light's kappa value of 0.46. Notably, six clinical manifestations achieved Light's kappa values > 0.80 . Therefore, the agreement varies from moderate to perfect.

Frequency of Reporting

The frequencies of the clinical manifestations are listed in Table 2. Coughing, dyspnea, fatigue, pain, and weight loss were recorded in more than 90% of the patients, with pain being the most frequently reported clinical manifestation (1169 occurrences). Constipation, diarrhea, hemoptysis, headache and hoarseness are recorded in less than 70% of patients, with hoarseness being the least frequent reported clinical manifestation (83 occurrences). Although fatigue is the most prevalent clinical manifestation affecting all but one patient, it has only 585 occurrences, with an average of 6.2 occurrences per patient.

Table 2 Frequency of Clinical Manifestations

Clinical Manifestations	No. of Patients (%)	No. of Occurrences (per Affected Patients)
Constipation	58 (61)	133 (2.3)
Coughing	87 (92)	470 (5.4)
Diarrhea	53 (56)	170 (3.2)
Dyspnea	93 (98)	899 (9.7)

(Continued)

Table 2 (Continued).

Clinical Manifestations	No. of Patients (%)	No. of Occurrences (per Affected Patients)
Fatigue	94 (99)	585 (6.2)
Hemoptysis	64 (67)	133 (2.1)
Headache	60 (63)	200 (3.3)
Hoarseness	49 (52)	83 (1.7)
Nausea	77 (81)	349 (4.5)
Pain	93 (98)	1169 (12.6)
Weight loss	92 (97)	459 (5.0)

Notes: The number of patients corresponds to the patients having at least one record. The “No. of occurrence” column represents the number of days with at least one record of the clinical manifestations. The values in parenthesis represents the average number of occurrences for patients with at least one record.

Abbreviation: No, number.

Danish National Patient Registry Compared to Patient Journals

Figure 2 shows a match of exclusively positive findings for the clinical manifestations. Overall, the precision is often high, whereas the recall tends to be low. The F1-score was low for nearly all clinical manifestations when aggregated by date, week, month, quarter, and year. When aggregating at the patient level, only dyspnea, hemoptysis, and pain had an F1-score above 0.25, with only dyspnea reaching above 0.5. This suggests a moderate level of reporting of these clinical manifestations at the patient level in the DNPR. Only dyspnea and pain reached an F1-score above 0.25 aggregating at the year level, and only dyspnea reached the threshold aggregating at the quarter level.

F1-scores at the patient, year, and quarter levels are listed in [Supplementary Tables 3, 4, and 5](#), respectively. Multiple clinical manifestations have an F1-score of 0, although specific SKS codes exist to report them. This indicates that most instances of clinical manifestations recorded by raters from patient journals did not appear in the DNPR.

Discussion

Main Results

This study compared data extracted from patient journals and the DNPR to evaluate how accurately the DNPR reflect clinical manifestations in lung cancer patients.

The high interrater agreement for most symptoms indicates that the manual journal data extraction was generally reliable, although the lower agreement for constipation ($\kappa = 0.46$) suggest that some clinical manifestations are more prone to subjective interpretation.

The concordance between the journal data and the DNPR was consistently low across all time intervals, highlighting the limitations of the DNPR, as an administrative tool, for clinical manifestation reporting. At the patient level, only severe clinical manifestations, such as dyspnea, hemoptysis, and pain reached an F1-score >0.25 , with only dyspnea reaching >0.5 , suggesting a moderate level of reporting of these clinical manifestations at the patient level in the DNPR. This pattern that the DNPR primarily reflects clinical manifestations that are clinically significant and likely to influence treatment decisions or resource use, while minor or moderate clinical manifestations stay underreported.

Strength and Limitations

This population-based study analyzed a randomly selected population of patients with lung cancer at the Department of Oncology in the NDR. Random selection limits selection bias.¹⁷ However, including patients from only one region could lead to a lack of generalizability of the results to Denmark and the rest of Scandinavia.¹⁰

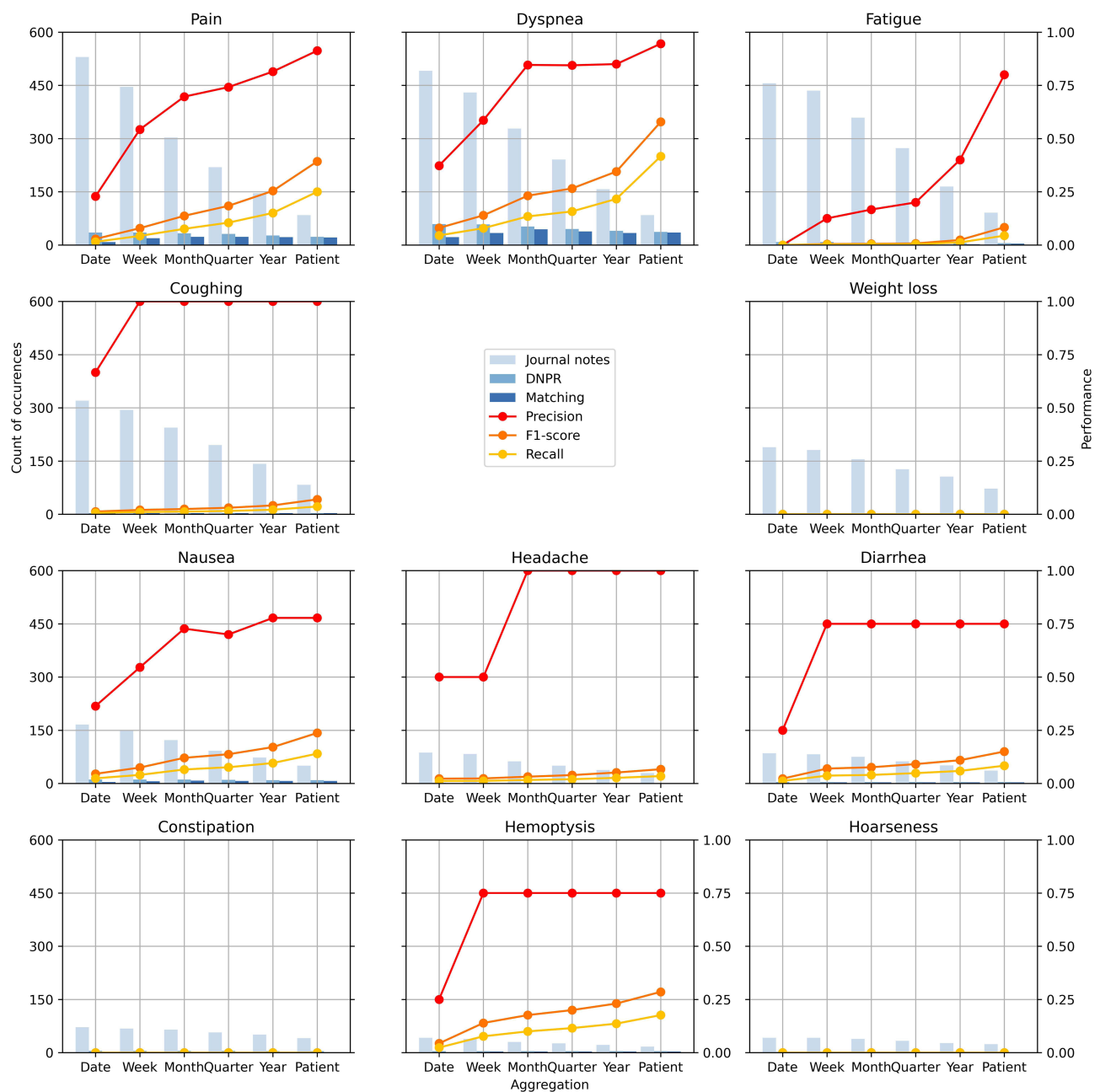


Figure 2 Count of occurrences in journal notes, the DNPR and matching between the two, and performance of precision, F1-score, and recall for pain, dyspnea, fatigue, coughing, weight loss, nausea, headache, diarrhea, constipation, hemoptysis, and hoarseness. Abbreviations: DNPR, Danish National Patient Registry.

Additionally, the study population comprised 95 patients, representing a relatively small cohort, that could hinder the generalizability of the results. The size of the cohort, provided the scope of the project, was nevertheless deemed sufficient to provide an indication of the concordance between journal data and the DNPR.

We decided to include patients based, not on their diagnosis date, but on their death date. The rationale was to have a complete picture of each patient trajectory, from diagnosis to death. This design choice could be seen as conditioning on future events but death being an ineluctable outcome for all patients, we considered it as acceptable since there was a limited patient selection on diagnosis date. Indeed, thanks to the long available history, having access to all patient data from 1991, we ensured a limited selection bias based on survival, allowing us to include patients with diverse survival

profiles, as illustrated in Table 1. Otherwise, it could have led to an overrepresentation of the most severe cases, reducing the generalizability of the findings.

The choice to scrutinize 12 patient journals to identify clinical manifestations can be considered arbitrary. Indeed, this was not based on a proper power calculation before the study, as it was intended to be exploratory. The assessments from these journals could possibly be biased owing to more thorough examinations compared with the remaining patients.

A limitation of this study could also be seen in the reliability of the reference standard. In this study, manually extracted patient journal data were considered the ground truth, leading to methodological challenges. First, the raters' professional background may have influenced data quality. Although supervised by an oncologist, students may be more prone to misinterpret clinical notes and the severity of the clinical manifestations, which could be the cause of the low interrater agreement for constipation (Light's kappa = 0.46). Second, the clinical manifestations reported by the patients were selectively recorded by clinicians, depending on their clinical relevance and significance. Thus, the patient journals themselves can be inherently incomplete and non-standardized leading to documentation bias. Third, the two data sources use different systems for describing clinical manifestations. The gradings based on CTCAE shows how severe the clinical manifestation is, while the DNPR codes only tell us if the clinical manifestation is present or not. In this study, any CTCAE grade above 0 was treated as equal to a DNPR entry, but this may be seen as an oversimplification.

Interrater Agreement

Interrater agreement analysis was conducted using Light's kappa to assess the similarity of ratings by the raters across 11 clinical manifestations in the patient journals of the four patients.

The interrater agreement varies among clinical manifestations, as shown in Figure 1. This variation could be attributed to several factors. First, some clinical manifestations are more difficult to grade because of an insufficient level of detail in patient journals. Second, the total number of clinical manifestations extracted varied among raters. Human error, either by overlooking a clinical manifestation or documenting one that is not present, may account for missing or added value. Therefore, Light's kappa was calculated without accounting for the different numbers of clinical manifestations noted by each rater.

When employing Light's kappa, an extension of Cohen's kappa, it is expected that the data will be on a nominal scale.¹⁸ However, only a few gradings adhered to this scale, while the majority were on an ordinal scale. In addition, the standardized interpretation of kappa values is criticized when used in health-related studies because it allows for a kappa of 0.41 to be acceptable.¹⁴ Therefore, it is suggested that a kappa value ≥ 0.60 is the acceptable standard in health-related studies.¹⁴ Thus, the only clinical manifestation with an unacceptable agreement was constipation (Light's kappa = 0.46). These findings indicate that the manually extracted data from patient journals are reliable. However, as the mean Light's kappa value for constipation is under the acceptable standard in health-related studies, manually extracted data on this cannot necessarily be considered the truth.

Concordance Between Patient Journals and the Danish National Patient Registry

This study explored whether the DNPR could serve as a source for clinical manifestations for decision support tools and future research. While the DNPR did not reflect the full spectrum of clinical manifestations documented in patient journals, it did reflect, in some cases, severe and clinically significant manifestations, such as dyspnea, hemoptysis, and pain. Dyspnea and pain were the most frequently reported and showed the highest concordance with journal data. These findings align with a study by Chowienczyk et al who identified dyspnea and cough as common presenting symptoms in lung cancer patients from 2000 to 2017.¹⁹ Although coughing was documented in most patient journals, its low F1-score indicates that it is less consistently recorded in the DNPR. In addition, Chowienczyk et al found that the presence of chest pain is dynamic and decreases over time.¹⁹ However, the current study generally considered pain beyond the chest pain.

It is important to interpret these findings in the light of the DNPR's primary purpose. As a primarily administrative tool, its coding practices prioritize diagnoses, treatments, and clinically significant events that influence health care resource allocation, rather than providing the detailed symptom documentation typical of clinical trial case forms.²⁰

Clinical Manifestations Report in Decision Support Systems

Clinical manifestations are crucial variables for assessing the quality of care and potentially predicting outcomes. While technically possible, the reporting of these data in the DNPR is only sparse, making them challenging to integrate in predictive models. Journal notes contain much more detailed information that could be used as input from predictive models leveraging Natural Language Processing (NLP) methods, notably to extract clinically relevant information.

Conclusion

This study found limited concordance between manually extracted data from patient journal notes and data from the DNPR. This highlights the limitations of using the DNPR to obtain reliable data for symptoms and side effects for lung cancer patients. Moderate concordance was nevertheless observed for dyspnea, hemoptysis, and pain at the patient level. This study points toward the potential for NLP to be applied for the extraction of clinical manifestations directly from journal notes to obtain more detailed data for epidemiological studies or to build machine learning-based decision support tools.

Disclosure

Dr Charles Vesteghem reports grants from Nordic Innovation, Innovation Fund Denmark, BETA.HEALTH, European Commission; personal fees from Interreg/European Union, outside the submitted work. Prof Martin Bøgsted reports grants from the Novo Nordisk Foundation and the European Commission. The authors declare no other conflicts of interest in this work.

References

1. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet*. 2021;398(10299):535–554. doi:10.1016/S0140-6736(21)00312-3
2. © International Agency for Research on Cancer. Age-standardized rate (nordic) per 100 000, incidence and mortality, both sexes. Available from: https://nordcan.iarc.fr/en/dataviz/trends?cancers=160&sexes=0&populations=208&age_start=0&age_end=17&types=0_1. Accessed May 17, 2025.
3. Kim J, Lee H, Huang BW. Lung Cancer: diagnosis, Treatment Principles, and Screening. *Am Fam Physician*. 2022;105(5):487–494.
4. Wallington M, Saxon EB, Bomb M, et al. 30-day mortality after systemic anticancer treatment for breast and lung cancer in england: a population-based, observational study. *Lancet Oncol*. 2016;17(9):1203–1216. doi:10.1016/S1470-2045(16)30383-7
5. Sundhedsstyrelsen. Pakkeforløb for lungekræft. 2024. Available from: <https://www.sst.dk/-/media/Udgivelser/2018/Lungekraeft/Pakkeforloeb-for-lungekraeft-jan-2025.ashx>. Accessed January 12, 2025.
6. Sengeløv L. Strålebehandling, bivirkninger - patienthåndbogen. Sundhed.dk Web site. Available from: <https://www.sundhed.dk/borger/patienthaandbogen/kraeft/sygdomme/behandlingsmetoder/straalebehandling-bivirkninger/>. Accessed April 10, 2024.
7. Sengeløv L. Kemoterapi - patienthåndbogen. Sundhed.dk Web site. Available from: <https://www.sundhed.dk/borger/patienthaandbogen/kraeft/sygdomme/behandlingsmetoder/kemoterapi/>. Accessed April 10, 2024.
8. Vesteghem C, Szejniuk WM, Brøndum RF, Falkmer UG, Azencott C, Bøgsted M. Dynamic risk prediction of 30-day mortality in patients with advanced lung cancer: comparing five machine learning approaches. *JCO Clin Cancer Inform*. 2022;6(6):2200054. doi:10.1200/CCI.22.00054
9. Schmidt M, Pedersen L, Sørensen HT. The danish civil registration system as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541. doi:10.1007/s10654-014-9930-3
10. Vesteghem C, Brøndum RF, Falkmer UG, Pottgård A, LØ P, Bøgsted M. High validity of the danish national patient registry for systemic anticancer treatment registration from 2009 to 2019. *Clin Epidemiol*. 2021;13:1085–1094. doi:10.2147/CLEP.S332776
11. U.S. Department of Health and Human Services. Common terminology criteria for adverse events (CTCAE). 2017.
12. Sundhedsdatastyrelsen. Classifications. sundhedsdatastyrelsen.dk Web site. Available from: https://sundhedsdatastyrelsen.dk/da/english/health_data_and_registers/classifications. Accessed August 5, 2024.
13. Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull*. 1971;76(76):365–377. doi:10.1037/h0031643
14. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276–282. doi:10.11613/BM.2012.031
15. Stata.com. Kappa — interrater agreement. Available from: <https://www.stata.com/manuals/rkappa.pdf>. Accessed April 10, 2024.
16. Chinchor N. MUC-4 evaluation metrics. *Proceedings of the 4th conference on Message understanding - MUC4 '92*. 1992:22–29. doi: 10.3115/1072064.1072067.
17. Buyse M. Statistics in oncology clinical trials. Phase III design: principles. Chinese Clinical Oncology Web site. Available from: <https://cco.amegroups.org/article/view/4672/html>. Accessed January 12, 2025.
18. Warrens MJ. Inequalities between multi-rater kappas. *Adv Data Anal Classif*. 2010;4(4):271–286. doi:10.1007/s11634-010-0073-4
19. Chowieńczyk S, Price S, Hamilton W. Changes in the presenting symptoms of lung cancer from 2000–2017: a serial cross-sectional study of observational records in UK primary care. *Br J Gen Pract*. 2020;70(692):e193–e199. doi:10.3399/bjgp20X708137
20. Sundhedsdatastyrelsen. Landspatientregisteret. www.sundhedsdatastyrelsen.dk Web site. Available from: <https://sundhedsdatastyrelsen.dk/data-og-registre/nationale-sundhedsregistre/landspatientregisteret>. Accessed December 7, 2025.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress

Taylor & Francis Group