

A Bayesian mixed effects and competitive risk approach for analysis of real-world treatment outcomes in small cell lung cancer

Luca Marzano, Asaf Dan, Adam S. Darwich, Luigi De Petris, Salomon Tendler, Rolf Lewenshon, Jayanth Raghothama, Sebastiaan Meijer

Submitted to: JMIR Cancer
on: September 13, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 31

 Multimedia Appendixes 32

 Multimedia Appendix 1..... 32

 Multimedia Appendix 2..... 32

A Bayesian mixed effects and competitive risk approach for analysis of real-world treatment outcomes in small cell lung cancer

Luca Marzano^{1*}; Asaf Dan^{2*}; Adam S. Darwich¹; Luigi De Petris²; Salomon Tendler³; Rolf Lewenshon²; Jayanth Raghothama¹; Sebastiaan Meijer¹

¹ KTH Royal Institute of Technology Stockholm SE

² Karolinska Institute Stockholm SE

³ Memorial Sloan Kettering Center New York US

*these authors contributed equally

Corresponding Author:

Adam S. Darwich

KTH Royal Institute of Technology

darwich@kth.se

Stockholm

SE

Abstract

Background: Small cell lung cancer (SCLC) is a challenging disease to treat due to rapid progression, development of chemo-resistance, and discrepancies in outcomes between real-world data and clinical trials. There is a lack of comprehensive analyses in other studies with regards to intermediate events and the treatment process, such as treatment decisions, progression of disease, and the occurrence of adverse events (AEs) over time.

Objective: The aim of this study was to apply advanced statistical methods on a longitudinal SCLC data set in order to identify factors of importance for the risk of AEs and for survival.

Methods: Treatment pathways of 421 SCLC patients collected from Karolinska University Hospital between 2016-2022 (Stockholm, Sweden) were analysed with data-driven modelling. The analysis focused on the impact of dose-adjustment on AEs, including neutropenia, by estimating odds ratios (OR) using Bayesian mixed-effects modelling. Covariate's effects on ECOG performance status (PS) deterioration and early discontinuation of chemotherapy with cause-specific hazard ratios (csHR) were explored using competitive risk models. This approach was applied to patient cohorts receiving combinatorial first line platinum/etoposide, and second line platinum/etoposide or platinum/irinotecan.

Results: At the end of the first line treatment, most patients exhibited tumour regression (n=167). Patients with neutropenia had longer overall survival (HR: 0.70 [0.53, 0.92]).

Higher etoposide dose levels were associated with subsequent occurrences of AEs (OR: 5.97 [1.41, 30.5]) and neutropenia (OR: 3.55 [1.03, 13.3]). Dose adjustment did not affect overall survival if the patient completed the four-dose regimen treatment. With regards to second-line therapy, fewer patients completed four treatment cycles and the most common reason of early discontinuation was tumour progression (n=72, 58%). Male patients (n=118) experienced fewer AEs and better first line treatment response compared to females (csHR: 0.51 [0.25, 0.90]). High-risk patients (defined as ECOG PS 2-3, or age > 75 years) with early discontinuation of therapy had survival outcomes similar to those who did not receive any therapy.

Conclusions: Our results indicate that first line therapies may benefit from more individualized dosing strategies. It would also be beneficial to assess the risk-benefit of treating specific subgroups, including patients receiving second line therapy. Real-world data proved beneficial for studying therapy response and risk-benefit of treating patient groups that are underrepresented in clinical trials.

(JMIR Preprints 13/09/2025:84042)

DOI: <https://doi.org/10.2196/preprints.84042>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

A Bayesian mixed effects and competitive risk approach for analysis of real-world treatment outcomes in small cell lung cancer

Luca Marzano ^{#a}, Asaf Dan ^{#b}, Adam S. Darwich ^{*a}, Luigi De Petris ^b, Salomon Tendler ^{bc}, Rolf Lewensohn ^b, Jayanth Raghothama ^a, Sebastiaan Meijer ^a

**Corresponding author: Adam S. Darwich, Hälsovägen 11 C
141 57 Huddinge (Sweden), darwich@kth.se*

Contributed equally

^aDivision of Health Informatics and Logistics, School of Engineering Sciences in Chemistry, Biotechnology and Health (CBH), KTH Royal Institute of Technology, Stockholm, Sweden

^bDept. of Oncology-Pathology, Karolinska Institutet and the Thoracic Oncology Center, Karolinska University hospital, Stockholm, Sweden

^cDepartment of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY

Abstract

Background: Small cell lung cancer (SCLC) is a challenging disease to treat due to rapid progression, development of chemo-resistance, and discrepancies in outcomes between real-world data and clinical trials. There is a lack of comprehensive analyses in other studies with regards to intermediate events and the treatment process, such as treatment decisions, progression of disease, and the occurrence of adverse events (AEs) over time. The aim of this study was to apply advanced statistical methods on a longitudinal SCLC data set in order to identify factors of importance for the risk of AEs and for survival.

Methods: Treatment pathways of 421 SCLC patients collected from Karolinska University Hospital between 2016-2022 (Stockholm, Sweden) were analysed with data-driven modelling. The analysis focused on the impact of dose-adjustment on AEs, including neutropenia, by estimating odds ratios (OR) using Bayesian mixed-effects modelling. Covariate's effects on ECOG performance status (PS) deterioration and early discontinuation of chemotherapy with cause-specific hazard ratios (csHR) were explored using competitive risk models. This approach was applied to patient cohorts receiving combinatorial first line platinum/ etoposide, and second line platinum/ etoposide or platinum/ irinotecan.

Results: At the end of the first line treatment, most patients exhibited tumour regression (n=167). Patients with neutropenia had longer overall survival (HR: 0.70 [0.53, 0.92]).

Higher etoposide dose levels were associated with subsequent occurrences of AEs (OR: 5.97 [1.41, 30.5]) and neutropenia (OR: 3.55 [1.03, 13.3]). Dose adjustment did not affect overall survival if the patient completed the four-dose regimen treatment. With regards to second-line therapy, fewer patients completed four treatment cycles and the most common reason of early discontinuation was

tumour progression (n=72, 58%). Male patients (n=118) experienced fewer AEs and better first line treatment response compared to females (csHR: 0.51 [0.25, 0.90]). High-risk patients (defined as ECOG PS 2-3, or age > 75 years) with early discontinuation of therapy had survival outcomes similar to those who did not receive any therapy.

Conclusions: Our results indicate that first line therapies may benefit from more individualized dosing strategies. It would also be beneficial to assess the risk-benefit of treating specific subgroups, including patients receiving second line therapy. Real-world data proved beneficial for studying therapy response and risk-benefit of treating patient groups that are underrepresented in clinical trials.

Introduction_

Small cell lung cancer (SCLC) is an aggressive and challenging disease to treat, mainly due to the lack of knowledge surrounding chemo-resistance mechanisms^{1,2}. Standard SCLC treatment consists of defined treatment lines, each comprising of four-six cycles of platinum-based chemotherapy (Cisplatin or Carboplatin) with Etoposide (PE) or Irinotecan (IP). For patients with cancer stages I–III, this is often combined with concurrent thoracic radiotherapy (RT). In stage IV cancer, immunotherapy with checkpoint inhibitors can be combined with chemotherapy, though with modest survival benefits^{3,4}. However, most SCLC patients experience relapse within 6 months after initial treatment, necessitating further therapy, though with limited survival⁵.

In the setting of disease relapse, the distinction between sensitive- and resistance-relapse plays a crucial role in predicting survival. Patients with sensitive disease, defined by relapse occurring more than 90 days after the completion of first-line platinum-based chemotherapy, might be re-challenged with the same chemotherapy regimen administered as first-line, and demonstrate significantly better survival outcomes compared to those with resistance disease, where relapse occurs within 90 days^{6,7}. In this latter case, the most common approach is to change to other chemotherapy agents, if the patient is still in acceptable general condition.

The rapid progression of SCLC often makes it difficult to manage⁸ which results in discrepancies in observed outcomes between clinical trials and real-world patients⁹. The knowledge gap between clinical trials and the real-world underrepresented patient populations can be overcome by treatment adjustments based on clinical experience, including choice of chemotherapy agents, dose adjustments, and end-of-life care^{8,10,11}. The focus of previous studies has mainly been on survival outcomes, often overlooking patients who received subsequent treatment lines after first-line chemotherapy. Further, there is a lack of studies that analyse events that occur over time in the treatment process^{8,12–14}.

Real-world data (RWD), here defined as healthcare data produced for the primary purpose of treatment in a clinical setting, offers an opportunity to retrospectively explore the impact of treatment decisions and pathways on patient outcomes given the variability in the real-world patient population^{15,16}. As such, RWD has the potential to inform the individualisation of treatments for specific subpopulations^{17,18}, such as patients with poor ECOG (Eastern Cooperative Oncology Group) performance status (PS), elderly patients, and patients with brain metastases who tend to be underrepresented in clinical trials⁹.

However, evidence generation from RWD is challenging because of complex clinical workflows and issues related to its secondary use, such as inherent biases in data collection and decision-making

processes, and non-observable effects^{19,20}. In SCLC treatment, the nuances of dose adjustments, planned treatment pauses, assessment of response, and early discontinuation of treatment due to severe toxicity or health deterioration, exemplify the intricacies involved^{21,22}.

The aim of this study is to describe the time-dynamics of patient trajectories in a clinical SCLC patient cohort, in order to identify factors of importance for dose-adjustments and survival outcomes. This approach necessitates an appropriate choice of data analysis²³. While artificial intelligence and machine learning have shown much promise for analysing RWD, issues of transparency and interpretability pose a potential risk of propagating biases from the data, including not captured information and operational confounders^{9,24}.

Here the focus was on selecting transparent modelling approaches^{24,25}, allowing for clinical contextualisation and interpretation^{9,20}. Then, the analysis consisted of a combination of Bayesian²⁶⁻²⁸ and mixed-effects models^{10,29,30} for studying dose adjustments and their impact on AEs (Bayesian logistic regression³¹) and the grade of toxicity (Bayesian cumulative link ordinal regression³¹). Competitive risk Fine-Gray regression³² was applied to compute cause-specific hazard ratios and predict the time-course of the therapy, ECOG PS deterioration and early discontinuation of treatment, defined as discontinuation before completing four planned chemotherapy cycles due to AEs, or radiological and clinical signs of tumour progression.

Materials and Methods

Collection of longitudinal data

The study received ethical approval by the Swedish Ethical Review Authority (approval number: 2023-07389-01). Patients still alive at the time of data collection signed a written informed consent. Research was performed in accordance with the Declaration of Helsinki. The study included a total of 421 patients diagnosed with SCLC at the Karolinska University Hospital in Stockholm, Sweden, between 2016 and 2022. Baseline patient characteristics such as age, sex, smoking status, TNM (Tumour, Node, Metastasis) stage (IASLC, International Association for the Study of Lung Cancer, 8th edition), Charlson Comorbidity Index (CCI), CNS metastasis, and previous cancer prior to SCLC were recorded. ECOG PS was recorded at baseline and at the start and cessation of each treatment line. Blood values, including haemoglobin, creatinine, sodium (Na), lactate dehydrogenase (LDH), C-reactive Protein (CRP), and albumin, were computed at baseline and at the start of each treatment line.

The specific drug regimen received by each patient was recorded. Body surface area (BSA) and estimated glomerular filtration rate (GFR) were recorded before start of treatments. These two variables were used to compute the dose of chemotherapy administered for each patient (*e.g.*, GFR for carboplatin dosing using Calvert formula³³, and BSA for etoposide and irinotecan, See Supplementary Table S1). The date and percentage of dose administered for each chemotherapy treatment were registered, including the reason for any modifications or adjustment to the administered doses, for the calculation of dose interval and intensity.

The reason for dose reduction was recorded. AEs were classified by type and grade of severity of toxicity (GT) according to the Common Terminology Criteria for AEs (CTCAE) version 5. In this work high toxicity was defined as AEs with CTCAE G 3-4³⁴.

Details on radiotherapy, including the treatment indication (curative, consolidating, as a separate

treatment line or palliative) and fractionation were reported. At the end of each treatment line, the tumoral response was assessed: complete response (CR), partial response (PR), stable disease (SD), or signs of tumor progression.

The overall survival (OS) was calculated from the date of diagnosis until the patient's death or the end of the study period, whichever came first. For each treatment line, Progression-free survival (PFS) was calculated from the date of the administration of the first dose, until signs of progression or death.

Analysis description

Several aspects of the treatment process were analysed for each treatment line, including: dose-adjustments due to AEs, changes in ECOG PS from the start to the end of the treatment line, and early treatment discontinuation (stopping treatment before 4 cycles of chemotherapy due to signs of tumour progression, toxicity, deteriorating medical condition, or unwillingness of the patient to receive the treatment) versus no – early discontinuation, defined as cessation of a treatment after 4 cycles of chemotherapy due to PR or SD, followed by a pause.

The analytical pipeline (Figure 1) can be summarised in the following steps:

- selection of clinical scenarios and patient subgroups,
- exploratory data analysis of the longitudinal cohort,
- multivariate analysis of dose adjustments,
- multivariate analysis of competitive risks,
- analysis of high-risk subgroups,
- aggregation and discussion of results.

Selection of clinical scenarios and patient subgroups. Prior to commencing the analysis, the collected data was analysed with a broad perspective to identify key subpopulations of interest. The cohort was explored in terms of baseline characteristics at the first line treatment decision, and the administered chemotherapy lines using summary statistics and Sankey Flows for the subsequent treatment lines. Treatment lines with relevant sample size were detected for the next analysis steps. Moreover, special subgroups of patients receiving treatment but with highly fragile conditions (underrepresented in trials or subjected to high risk of poor survival outcomes) were selected.

Exploratory analysis of the longitudinal cohort. CTCAE G3-4 AEs, ECOG PS change at the end of treatment, and early discontinuation of therapy were assessed by exploring absolute count and relative percentages. Hazard ratios (HRs) and Kaplan Meier curves were computed to estimate the impact on survival outcomes³⁵. The cumulative incidence-time curves of dose-reduction due to AEs, separating patients with good therapy response and the ones with unplanned stops, were analysed³⁶. Moreover, the survival outcomes were estimated by stratifying patients by those receiving four doses (full course) without dose reduction, patients that received four cycles but with dose reduction, and patients that received less than four cycles.

Multivariate analysis of dose adjustments. For the dose adjustment analysis, the association between dose reductions caused by an AE and the grade of toxicity was explored. This was done by training two Bayesian mixed-effect regression models: a binary logistic regression for the adverse events incidence^{31,37}, and a cumulative link regression for the grade of toxicity modelled as an ordinal outcome^{31,38}. The longitudinal variables (dose level, dose interval, and their mutual interaction) were analysed with respect to the key baseline characteristics. The Bayesian logistic regression can be summarised as follows:

$$Y_{AE} = \log \frac{P(AE_{i+1})}{P(GT_{i+1} \leq j)}$$

Where $P(AE_{i+1})$ is the probability of a dose reduction due to adverse event at dose $i+1$ caused by the i :th dose D_i , with interval Δt_i between the doses, x baseline covariates and ϵ random effects. The cumulative link ordinal Bayesian regression can be summarised as follows:

$$Y_i = \log \left(\frac{P(GT_{i+1} \leq j)}{1 - P(GT_{i+1} \leq j)} \right) = b_{0j} + b_D D_i + b_{\Delta t} \Delta t_i + b_{D\Delta t} D_i \Delta t_i + \sum_{\text{baselines}} b_x x + \sum_{\text{random}} \epsilon$$

With similar notation for the Bayesian logistic regression, where $P(GT_{i+1} \leq j)$ is the probability of grade of toxicity lower than or equal to j (with $j = 0, \dots, 4$), and b_{0j} is the baseline intercept for the j :th level. The Bayesian models were trained running four independent Markov chains Monte Carlo (MCMC) with 2,500 iterations for each chain. The odds ratios (exponential of the coefficients b) with the 95% confidence level were estimated for all the covariates. The random effects were estimated for the individual patients, administrated doses and the dose interval. This allowed for individual differences and the time-dependent nature of the dose variables to be considered.

Multivariate analysis of competitive risks. For the variation of ECOG PS at the end of the treatment line and the early discontinuation of the therapy, a competitive risk time-to-event model was developed. Fine-Gray competitive risk regression was used with the outcomes³¹, the progression free survival (PFS) associated to the investigated treatment line and the event of interest to estimate the distributed Hazard ratios of the covariates. The model can be summarised as follows:

$$h_j = h_0 \exp(b_1 x_1 + \dots + b_m x_m)$$

Where, h_j is the cause-specific hazard of the event of interest, h_0 the baseline hazard, and the exponential of the m :th coefficient b the distributed hazard ratio of the m :th covariate x . The cause-specific hazard ratios were computed for ECOG PS deterioration/improvement (in competition with unchanged PS or improvement and deterioration), and the early discontinuation of the therapy in competition with planned stop of treatment due to PR or SD. Here we indicate the OS or PFS Cox hazard ratios with HR, csHR for the cause-specific hazard ratios computed with the Fine-Gray model, and OR for the odds ratios obtained by the Bayesian mixed-effects models.

High-risk subgroups analysis. Multivariate analysis was performed for the subgroups of patients with detected fragile conditions receiving treatment. The OS of these patients was compared to the OS of patients with similar characteristics that did not receive chemotherapy. This allowed the comparison of survival outcomes of patients with poor response with the ones who did not receive chemotherapy, thus comparing the benefits and risks of the received treatment.

Aggregation and discussion of results. The considerations obtained during the exploratory data analysis are discussed together with the results obtained from the multivariate analysis. This allows to create a connection between covariate effects on the longitudinal outcomes that can impact on the OS or treatment decision. Moreover, the high-risk subgroups patients are compared with the ones that did not received treatment stratified by the significant covariates or longitudinal outcome discovered during the multivariate subgroup analysis to assess the risk-benefit of receiving treatment. Finally, the overall outcomes are discussed with the clinical experts regarding the potential clinical implications.

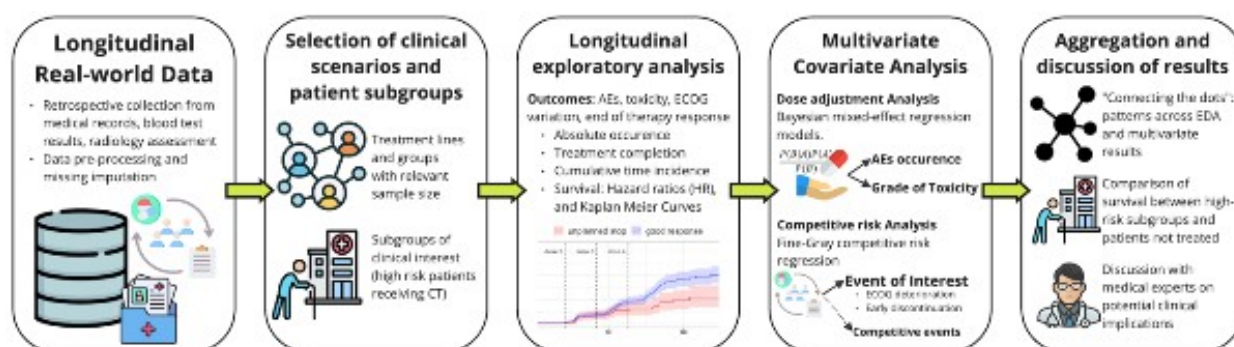


Figure 1. Analytical pipeline description.

Table1. Summary of the analysed cohort. 1L: first line, 2L: second line, PE: platinum etoposide, IP platinum Irinotecan, ECOG PS: performance status, CCI: Charles Comorbidity Index, CNS: brain metastasis, LDH: lactate hydrogenasis, CRP: c-reactive protein, Na: sodium, BSA: body surface area, GFR: estimated glomerular filtration rate, AEs: adverse events, PFS: progression free survival, OS: overall survival. SD: standard deviation. (*): log transformation applied during the analysis to handle high skewness.

Covariates		1L	2L	
Treatment process context	Variables	PE (N=289)	PE (N=69)	IP (N=56)
Baselines	Age (years)			
	<75ys	190 (65.7%)	51 (73.9%)	47 (83.9%)
	>=75ys	99 (34.3%)	18 (26.1%)	9 (16.1%)
	Sex			
	Female	171 (59.17%)	34 (49.3%)	29 (51.8%)
	Male	118 (40.83%)	35 (50.7%)	27 (48.2%)
	ECOG PS baseline			
	0	58 (20.1%)	16 (23.2%)	15 (26.8%)
	1	117 (40.5%)	35 (50.7%)	29 (51.8%)
	2	75 (26.0%)	13 (18.8%)	9 (16.1%)
	3	39 (13.5%)	5 (7.2%)	3 (5.4%)
	CCI			
	Mean (SD)	8.11 (2.30)	7.22 (2.18)	8.07 (1.87)
	Median [Min, Max]	9.00 [2.00, 15.0]	8.00 [3.00, 12.0]	8.00 [3.00, 12.0]
	Missing	4 (1.4%)	2 (2.9%)	2 (3.6%)
	CNS			
	no	247 (85.5%)	66 (95.7%)	46 (82.1%)
	yes	42 (14.5%)	3 (4.3%)	10 (17.9%)
	previous cancer			
	no	158 (54.7%)	50 (72.5%)	44 (78.6%)
	yes	60 (20.8%)	19 (27.5%)	12 (21.4%)
	Missing	71 (24.6%)		

TNM stage	TNM stage			
	IIIA-B	61 (21.1%)	19 (27.5%)	8 (14.3%)
	IIIC	19 (6.6%)	6 (8.7%)	1 (1.8%)
	IVA	74 (25.6%)	20 (29.0%)	15 (26.8%)
	IVB	135 (46.7%)	24 (34.8%)	32 (57.1%)
	TNM-T			
	T1-2	70 (24.2%)	16 (23.2%)	9 (16.1%)
	T3	34 (11.8%)	9 (13.0%)	8 (14.3%)
	T4	185 (64.0%)	44 (63.8%)	39 (69.6%)
	TNM-N			
	N0	20 (6.9%)	9 (13.0%)	6 (10.7%)
	N1	12 (4.2%)	3 (4.3%)	2 (3.6%)
	N2	101 (34.9%)	24 (34.8%)	17 (30.4%)
	N3	155 (53.6%)	33 (47.8%)	31 (55.4%)
	Missing	1 (0.3%)		
	TNM-M			
	M0	80 (27.7%)	25 (36.2%)	9 (16.1%)
	M1a	37 (12.8%)	13 (18.8%)	6 (10.7%)
	M1b	37 (12.8%)	7 (10.1%)	9 (16.1%)
	M1c	135 (46.7%)	24 (34.8%)	32 (57.1%)
blood values and weight at the start of the treatment after MissingForrest Imputation	Hb (g/L)			
	Mean (SD)	130 (16.5)	127 (19.5)	122 (14.1)
	Median [Min, Max]	130 [87.0, 170]	128 [14.0, 156]	120 [100, 156]
	Crp* (mg/L)			
	Mean (SD)	34.7 (49.6)	10.3 (11.8)	21.8 (38.9)
	Median [Min, Max]	15.0 [1.00, 331]	7.00 [1.00, 69.0]	10.0 [1.00, 267]
	Na (mmol/L)			
	Mean (SD)	137 (4.97)	138 (4.93)	137 (4.41)
	Median [Min, Max]	137 [112, 148]	139 [127, 151]	137 [120, 146]
	Ldh* (μkat/L)			
	Mean (SD)	26.1 (44.4)	23.4 (46.9)	22.7 (31.5)
	Median [Min, Max]	9.30 [2.00, 607]	10.8 [2.60, 377]	11.2 [3.00, 193]
	Creatinine* (μmol/L)			
	Mean (SD)	68.6 (22.9)	71.7 (21.9)	68.9 (20.3)
	Median [Min, Max]	66.0 [26.0, 204]	69.0 [38.0, 139]	65.0 [32.0, 123]
	albumin* (g/L)			
	Mean (SD)	32.7 (14.3)	34.8 (4.33)	34.1 (7.62)
	Median [Min, Max]	32.0 [3.00, 250]	35.0 [24.0, 43.0]	33.0 [25.0, 82.0]
	neutrophils count* (10⁹/L)			
	Mean (SD)	34.8 (44.5)	5.92 (2.66)	6.84 (3.66)

	Median [Min, Max]	19.2 [1.80, 514]	5.59 [0.500, 14.5]	6.23 [0.900, 26.4]
	weight (kg)			
	Mean (SD)	70.6 (15.3)	69.8 (12.5)	75.8 (16.7)
	Median [Min, Max]	68.5 [37.0, 125]	70.0 [45.0, 103]	78.0 [48.0, 113]
	Missing	3 (1.0%)	2 (2.9%)	3 (5.4%)
Dose calculation	BSA (m2)			
	Mean (SD)	1.79 (0.213)	1.84 (0.383)	1.84 (0.249)
	Median [Min, Max]	1.78 [1.18, 2.37]	1.81 [1.31, 4.53]	1.86 [1.09, 2.31]
	Missing	4 (1.4%)	2 (2.9%)	1 (1.8%)
	GFR (ml/min/1.7)			
	Mean (SD)	78.6 (28.7)	78.2 (25.4)	86.4 (41.0)
	Median [Min, Max]	75.0 [8.00, 235]	75.0 [30.0, 157]	80.0 [36.0, 299]
	Missing	11 (3.8%)	1 (1.4%)	2 (3.6%)
	Etoposide dose administration			
	ten days	69 (23.9%)	9 (13.0%)	-
	three days	220 (76.1%)	60 (87.0%)	-
Adverse effects dose reduction	Neutropenia			
	No	205 (70.9%)	43 (62.3%)	53 (94.6%)
	Yes	84 (29.1%)	26 (37.7%)	3 (5.4%)
	Other AEs			
	No	225 (77.9%)	52 (75.4%)	33 (58.9%)
	Yes	64 (22.1%)	17 (24.6%)	23 (41.1%)
	CTCAE G3-4			
	No	185 (64.0%)	39 (56.5%)	38 (67.9%)
	Yes	104 (36.0%)	30 (43.5%)	18 (32.1%)
End of treatment reported response	ECOG PS variation			
	stable	87 (30.1%)	36 (52.2%)	28 (50.0%)
	deterioration	141 (48.8%)	21 (30.4%)	17 (30.4%)
	improvement	61 (21.1%)	12 (17.4%)	11 (19.6%)
	treatment line response			
	Early discontinuation	110 (38.1%)	36 (52.2%)	37 (66.1%)
Radiotherapy (RT)	CR, PR, or SD	179 (61.9%)	33 (47.8%)	19 (33.9%)
	treatment intention			
	No RT	172 (59.5%)	47 (56.0%)	50 (70.4%)
	curative	45 (15.6%)	0 (0%)	0 (0%)
	consolidating	39 (13.5%)	4 (4.8%)	0 (0%)
	palliative	33 (11.4%)	30 (35.7%)	19 (26.8%)
Survival outcomes	PCI	-	3 (3.6%)	2 (2.8%)
	PFS (days)			
	Mean (SD)	185 (174)	157 (168)	127 (93.0)
	Median [Min, Max]	163 [0, 514]	125 [0, 1170]	110 [9.00, 26.4]

		1980]		473]
	Missing	43 (14.9%)	8 (11.6%)	4 (7.1%)
	OS (days)			
	Mean (SD)	311 (303)	607 (391)	355 (157)
	Median [Min, Max]	245 [11.0, 2310]	516 [140, 2310]	318 [144, 911]
	Status at the last reported information			
	dead	262 (90.7%)	63 (91.3%)	54 (96.4%)
	alive (censored)	27 (9.3%)	6 (8.7%)	2 (3.6%)

Results

Selection of clinical scenarios and patient subgroups

The cohorts taken into consideration were the subjects receiving platinum etoposide (PE) as first line therapy, and the patients receiving either rechallenge with PE, or platinum and irinotecan (IP) as second line therapy. The analysis included stage III-IV patients and excluded patients with ECOG PS of 4 (n=29). **Supplementary Figure S1** reports the inclusion/exclusion flowchart of the cohort. In terms of toxicity, only neutropenia was analysed in detail in this study since it was the most commonly reported treatment-related AE (See Supplementary Table S2). Supplementary Figures S1-4 show Sankey flows of the subsequent treatment lines and the dose adjustments.

The identified high-risk subgroups included the first line patients receiving PE with i) ECOG PS 2-3 (n=114) ii), with CNS metastases (n=42) iii), over the age of 75 years (n=99) and iv), with stage IVB (n=135). These subgroups were selected since they are known to be either underrepresented in RCTs or are associated with relatively poor OS compared to the general population⁹. The age threshold was chosen to define a sharp subgroup of elderly patients from the cohort. The multivariate analysis was not performed for patients with CNS metastases due to wide confidence intervals of ORs and csHR caused by the small sample size of the subgroup (n=42).

For some individuals, blood values and lab tests were missing (neutrophil counts 44.9%, LDH 22.9%, and CRP 18.3%). The percentages of missing values for neutrophil counts at the start of the treatment lines were as follows: 50.9% in the first line of platinum etoposide, 34.8% in the second line platinum etoposide, and 39.3% in the second line of platinum irinotecan. The high percentage of missing neutrophil values was due to these not being routinely measured at baseline before the initiation of chemotherapy treatment. Therefore, during the multivariate analyses, for each treatment line analysis, values were imputed using the machine learning algorithm MissForest³⁹. MissForest was chosen since it is particularly effective for datasets with mixed-type variables and complex interactions, maintaining low imputation errors and preserving relationships among variables even with high percentages of missing information⁴⁰.

A summary of the analysed cohorts for the first and the second line is reported in **Table 1**.

Exploratory survival analysis

In the first line setting, chemotherapy doses were reduced in 127 out of 289 patients (43.9%), mainly

due to any grade of neutropenia ($n=89$) and AE G3-4 ($n=104$) (Table 1). Similar figures were observed for the second line setting (PE $n=69$, IP $n=56$), with neutropenia mainly associated with PE treatment ($n=26$). At the end of the first line, $n=179$ (out of which $n=167$ PR) patients did not experience early discontinuation of treatment. Early discontinuation of therapy happened in 110 cases, due to clinical ($n=48$) or radiological ($n=23$) signs of PD. In the second line, early discontinuation ($n=72$) was more common than completion of treatment ($n=52$).

Figure 2 shows the cumulative incidences of dose adjustment due to neutropenia. The incidence was higher for patients in both lines who achieved treatment completion without premature discontinuation. In the first line scenario, occurrences of neutropenia and other AEs were associated with more favourable OS (HR: 0.70 [0.53, 0.92] and HR: 0.72 [0.57, 0.95] respectively). In the second line, only neutropenia was associated with better OS (HR: 0.51 [0.33, 0.80]). OS of patients completing four cycles of first-line treatment was not affected by dose reduction due to AEs (HR: 0.85 [0.59, 1.23], reference: patients completing four cycles without dose reduction) (Figure 3).

Patients with early discontinuation of therapy in the first line setting, showed an expected markedly worse OS compared to patients who completed 4 cycles chemotherapy (HR: 4.27 [3.27, 5.57]). Similarly, worse OS was observed for patients with early discontinuation during second line therapy (HR: 2.21 [1.50, 3.27]). Moreover, ECOG PS deterioration was significantly associated with a poorer prognosis, even if this was only seen in the first line setting (HR: 2.05 [1.54, 2.72], reference: stable ECOG PS). These patients resulted in early discontinuation of therapy from which $n=45$ resulted with PFS less than 3 months (median: 61, interquartile range [43, 68] days).

Patients receiving PE as second line completed the four-dose cycle more often than the IP patients (73.36% vs 41.07%, respectively). Interestingly, the PFS of PE patients was similar compared to patients receiving IP (HR: 0.84 [0.57, 1.23]). However, as expected, second line PE patients showed better OS (HR: 0.51 [0.33, 0.78] compared to the ones receiving IP, when overall survival was computed from the start of the second line treatment.

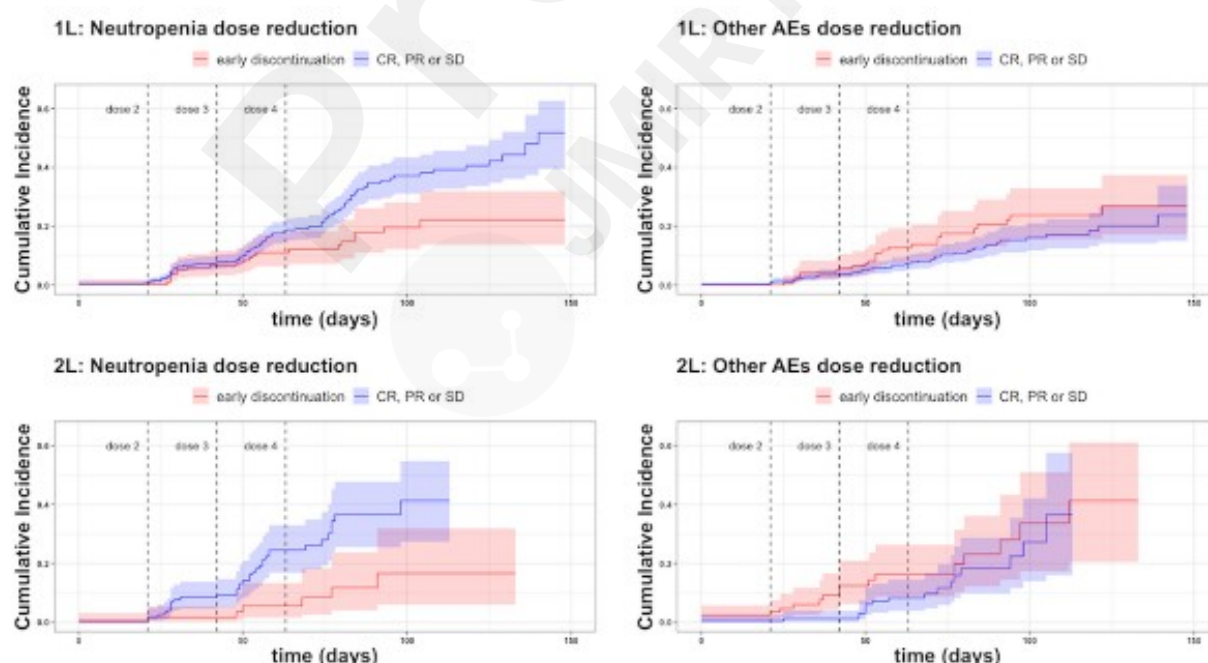


Figure 2. Cumulative incidences of dose reduction (%) due to neutropenia and other adverse events (AEs) for the first and the second line therapy. The curves are stratified by early discontinuation of chemotherapy (red

curves) and therapy response (CR complete, PR partial, SD stable disease) (blue curves). Dotted lines correspond to the three-week intervals corresponding to planned administration of the second, third and fourth doses.

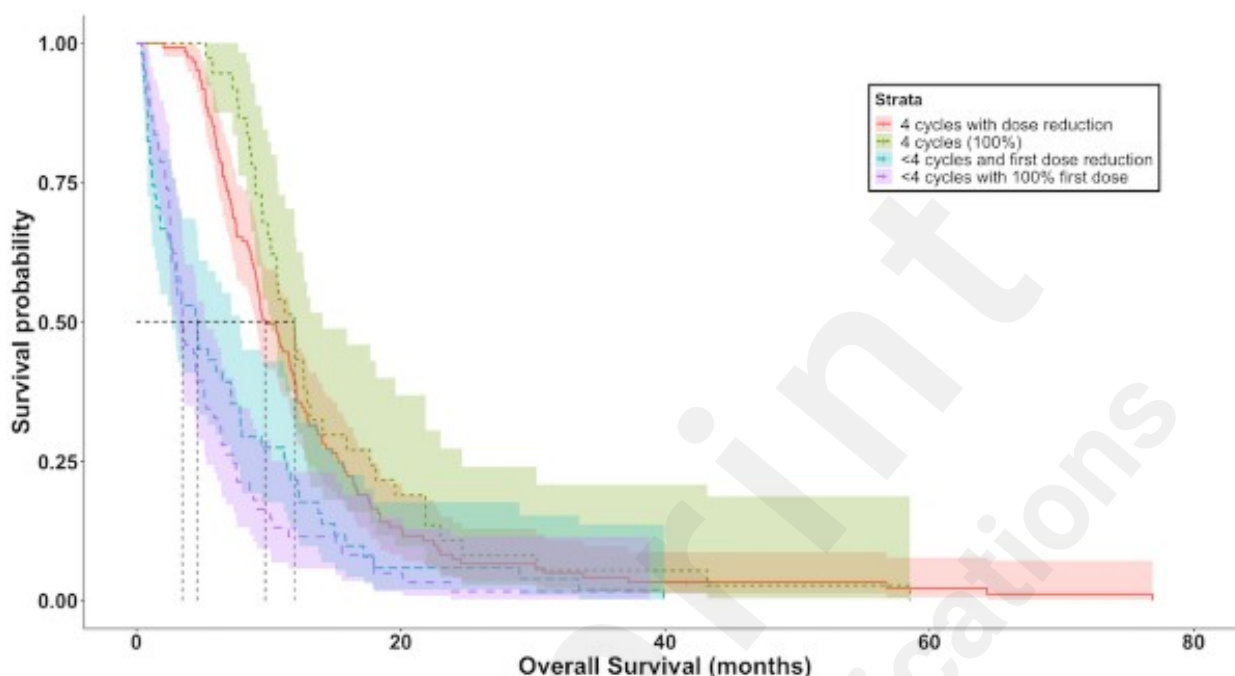


Figure 3. Kaplan Meier curves for overall survival of patients receiving first line treatment stratified by administered doses. Patients receiving less than four doses experienced early discontinuation of the therapy.

Multivariate analysis

First line therapy PE

Tables 2 and 3 detail the results of the dose adjustment and the competitive risk analysis for the first line therapy. All the longitudinal variables presented significant variability (See **Supplementary Table S3**). Inter-individual variability was higher than inter-occasion variability.

The results show a strong association between higher etoposide dose levels and the subsequent occurrence of AEs (OR: 5.97 [1.41, 30.5]), specifically neutropenia (OR: 3.55 [1.03, 13.3]). Early discontinuation of therapy was associated to higher LDH levels (csHR 1.44 [1.14, 1.82]), baseline ECOG PS 2 (csHR: 2.93 [1.31, 6.52], baseline ECOG PS 3 (csHR 4.15 [1.68, 10.2]).

Males showed less probability of early discontinuation of the first treatment line compared to females (csHR: 0.51 [0.29, 0.90]). As did subgroups of patients with M0 and those receiving consolidating radiotherapy (n=29) (Table 3). Similarly, patients with M1a status regardless of T and N stage were associated with ECOG PS improvement (Table 3).

Table2. Bayesian mixed-effects analysis for the first line platinum etoposide patients. OR: Odds ratio, C.I.: confidence interval, ECOG PS: performance status, CCI: Charles Comorbidity Index, CNS: brain metastasis, ldh: lactate hydrogenasis, crp: c-reactive protein, Na: sodium, BSA: body surface area, GFR: estimated glomerular filtration rate. (*) interaction between two covariates

Covariates	Adverse effects		Neutropenia Grade	
	OR	95% C.I.	OR	95% C.I.
dose interval	1.00	0.94, 1.05	0.99	0.95, 1.03
dose platinum	0.60	0.07, 4.33	0.56	0.14, 2.18
dose etoposide	5.97	1.41, 30.5	3.55	1.03, 13.3
dose platinum*dose interval	1.00	0.97, 1.03	1.01	0.98, 1.05
<i>dose etoposide*dose interval</i>	1.02	0.99, 1.06	<i>1.04</i>	<i>1.00, 1.08</i>
etoposide dose load				
Three days	-	-	-	-
Ten days	1.69	0.31, 9.23	5.03	1.25, 23.9
weight	0.28	0.03, 1.88	1.52	0.30, 7.47
gfr	3.14	0.47, 24.8	1.59	0.47, 5.39
BSA	0.53	0.04, 5.95	0.16	0.02, 1.12
hb	0.77	0.37, 1.50	0.88	0.52, 1.46
creatinine	3.98	1.45, 12.5	1.32	0.67, 2.56
crp	0.83	0.40, 1.76	0.99	0.57, 1.72
na	0.97	0.50, 1.91	1.11	0.68, 1.81
ldh	1.59	0.78, 3.41	1.56	0.91, 2.82
alb	0.61	0.30, 1.20	0.83	0.48, 1.40
Neutrophil counts	0.39	0.16, 0.82	0.67	0.36, 1.21
age				
>=75ys	-	-	-	-
<75ys	0.92	0.20, 4.42	0.77	0.26, 2.32
gender				
Female	-	-	-	-
Male	0.27	0.05, 1.39	1.96	0.47, 8.81
ECOG PS baseline				
0	-	-	-	-
1	0.34	0.07, 1.56	0.89	0.31, 2.61
2	0.49	0.07, 3.22	1.14	0.29, 4.71
3	0.45	0.04, 4.05	4.24	0.66, 36.5
TNM-T				
T4	-	-	-	-
T1-2	1.87	0.44, 8.35	0.68	0.23, 1.97
T3	0.50	0.07, 3.54	0.73	0.19, 3.04
TNM-N				
N3	-	-	-	-
N0	0.84	0.07, 10.1	1.92	0.30, 12.2
N1	5.65	0.17, 190	37.4	2.46, 721

N2	0.97	0.25, 3.86	1.03	0.37, 2.61
TNM-M				
M1c	-	-	-	-
M0	6.01	0.72, 61.9	1.47	0.31, 6.97
M1a	2.95	0.45, 21.2	1.14	0.30, 4.64
M1b	3.22	0.43, 25.6	0.47	0.10, 2.28
smoking status				
Smoker	-	-	-	-
Previous smoker	0.92	0.24, 3.49	1.11	0.39, 3.21
Non-smoker	0.62	0.02, 18.5	1.05	0.07, 14.8
CCI	1.49	0.55, 3.92	1.26	0.62, 2.63
CNS				
no	-	-	-	-
yes	0.70	0.10, 4.85	0.86	0.21, 3.63
previous cancer				
no	-	-	-	-
yes	0.66	0.13, 2.84	0.58	0.20, 1.57

Table 3. Competitive risk analysis results for the first line treatment platinum etoposide. csHR: cause specific Hazard Ratio, C.I.: confidence interval, ECOG PS: performance status, CCI: Charles Comorbidity Index, CNS: brain metastasis, ldh: lactate hydrogenasis, crp: c-reactive protein, Na: sodium, BSA: body surface area, GFR: estimated glomerular filtration rate, RT: radiotherapy, cur: curative, cons: consolidating, pall: palliative.

Covariates	Early discontinuation of treatment			ECOG PS Deterioration			ECOG PS Improvement		
	N	csHR	95% C.I.	N	csHR	95% C.I.	csHR	95% C.I.	

age								
<75ys	149	—	—	140	—	—	—	—
>=75ys	82	1.01	0.55, 1.84	64	1.57	0.75, 3.27	0.43	0.14, 1.33
gender								
Female	131	—	—	111	—	—	—	—
Male	100	0.51	0.29, 0.90	93	0.89	0.41, 1.96	0.52	0.13, 2.04
ECOG baseline	PS							
0	49	—	—	-	-	-	-	-
1	94	1.08	0.53, 2.19	-	-	-	-	-
2	61	2.93	1.31, 6.52	-	-	-	-	-
3	27	4.15	1.68, 10.2	-	-	-	-	-
CCI	231	1.10	0.92, 1.32	204	0.98	0.80, 1.22	1.09	0.83, 1.45
previous cancer								
no	181	—	—	159	—	—	—	—
yes	50	0.75	0.38, 1.50	45	0.77	0.34, 1.71	1.34	0.40, 4.46
CNS								
no	196	—	—	176	—	—	—	—
yes	35	0.83	0.40, 1.75	28	0.64	0.24, 1.71	1.50	0.34, 6.50
TNM-T								
T4	149	—	—	130	—	—	—	—
T1-2	52	1.55	0.82, 2.93	47	1.33	0.67, 2.62	0.79	0.29, 2.11
T3	30	0.78	0.42, 1.45	27	0.96	0.50, 1.85	0.54	0.13, 2.20
TNM-N								
N3	129	—	—	113	—	—	—	—
N0	17	0.17	0.04, 0.70	15	0.73	0.27, 1.98	0.48	0.13, 1.81
N1	8	1.05	0.36, 3.08	6	0.13	0.04, 0.49	0.00	0.00, 0.00
N2	77	0.80	0.47, 1.36	70	0.77	0.40, 1.51	0.59	0.22, 1.54
TNM-M								
M1c	119	—	—	102	—	—	—	—
M0	52	1.52	0.53, 4.41	47	0.49	0.11, 2.14	1.07	0.17, 6.67
M1a	29	1.25	0.58, 2.69	25	0.55	0.18, 1.69	3.05	1.04, 8.93
M1b	31	1.03	0.53, 1.97	30	1.21	0.54, 2.72	4.26	0.81, 22.5
hb	231	1.01	0.99, 1.03	204	1.00	0.98,	0.99	0.97, 1.02

ldh	231	1.44	1.14, 1.82	204	1.19	1.01 0.94, 1.49	0.98	0.58, 1.64
crp	231	1.13	0.92, 1.38	204	1.02	0.80, 1.31	1.22	0.89, 1.68
Neutrophil counts	231	0.99	0.73, 1.35	204	1.40	1.06, 1.85	0.51	0.32, 0.82
na	231	1.02	0.97, 1.07	204	1.01	0.95, 1.07	1.05	0.96, 1.15
creatinine	231	2.40	0.84, 6.83	204	2.54	0.68, 9.48	0.80	0.09, 7.39
weight	231	0.99	0.95, 1.03	204	1.06	1.00, 1.12	0.93	0.85, 1.02
gfr	231	1.00	0.99, 1.02	204	1.00	0.99, 1.01	0.99	0.97, 1.01
body area	231	2.97	0.14, 62.5	204	0.01	0.00, 1.58	144	0.07, 286,657
etoposide dose load								
Ten days	57		—	51	—	—	—	—
Three days	174	0.39	0.22, 0.69	153	0.72	0.35, 1.48	3.65	0.91, 14.7
RT								
No RT or pall.	175	—	—	151	—	—	—	—
cur	24	0.41	0.14, 1.24	22	1.35	0.42, 4.28	0.31	0.06, 1.47
cons	32	0.20	0.06, 0.69	31	0.50	0.16, 1.50	0.53	0.24, 1.17

Second line therapy PE-IP

Results of the multivariate analysis for the second-line treatment are reported in Supplementary Tables S4 and S5. Due to the large confidence intervals of the ORs, the ORs are reported on a log scale.

A higher dose of platinum was associated with a higher incidence of AEs and a higher risk of neutropenia. Patients receiving IP experienced fewer AEs. Male patients were less likely to experience AEs and neutropenia. However, male patients were associated with worse PFS (HR: 1.62 [1.10, 2.38]) due to early discontinuation of the second line regardless PE or IP (n=44). Higher sodium levels were instead associated with lower risk of neutropenia and ECOG PS deterioration. Patients with TNM-M M1a had a higher probability of ECOG PS deterioration and more AEs. CNS metastasis was linked to increased risk for ECOG PS deterioration and early discontinuation of therapy.

High-risk subpopulation analysis: to treat or not to treat?

In the Supplementary Tables S6-11 we report the results of the multivariate analysis for the high-risk subgroups. Table 4 summarises the significant covariates for each subgroup. Similarly to the second

line therapy analysis, the ORs were reported in log scale due to the wide confidence intervals.

Results observed in the subgroup analyses that aligned with the overall cohort were the association of LDH and ECOG PS 2-3 with early discontinuation of therapy (all subgroups), dose etoposide and increased risk of AEs and neutropenia (IVB and age ≥ 75 ys), lower AE risk for male patients (age ≥ 75 ys). For patients with ECOG PS 2-3 higher dose of platinum was associated with an increased risk of AEs and neutropenia.

The comparison with patients that did not receive chemotherapy showed that, even though AEs were associated with longer survival, the hazard ratios of patients with ECOG PS 2-3 or age ≥ 75 ys were closer to the non-treated subgroup rather than to the patients who experienced a good response to treatment (Supplementary Table S12). We observed that when further stratifying the high-risk subgroups using the significant variables reported in Table 4, the subgroups showed similar survival to the same group of patients that did not receive treatment. For example, Figure 4 shows the Kaplan Meier curves of the comparative analysis for patients with ECOG PS 2-3 and age ≥ 75 years.

Table 4. Summary of statistically significant covariates in the multivariate analysis for high-risk subgroups. (+): OR or csHR with a direct effect on the studied outcome, (-): OR or csHR with an inverse effect on the studied outcome, ECOG PS: ECOG performance status, CCI: Charles Comorbidity Index, ldh: lactate hydrogenasis, crp: c-reactive protein, BSA: body surface area. () interaction between two covariates*

High risk Subgroup	Adverse Effects occurrence	Neutropenia	ECOG PS deterioration	ECOG PS improvement	Early Discontinuation
TNM-IVB	Dose etoposide (+) dose*interval etoposide (+) albumin (-) Neutrophil counts (-)	Dose*interval etoposide(+) BSA (-)	Neutrophil counts(+)	Neutrophil counts(-) LDH(+)	LDH(+) Creatinine (+)
ECOG PS 2-3	dose*interval platinum (+)	dose*interval platinum (+)	Neutrophil counts (+)	Neutrophil counts (-) LDH (+)	etoposide 10-days regime (+) TNM-N0 (-) TNM-M0 (+) HB (+) LDH (+) CRP (+) CCI (+)
AGE ≥ 75	LDH (+) Albumin (-), Neutrophil counts (-) Male (-) ECOG 3 (-) TNM-T3 (-) TNM-N2 (+)	Dose*interval etoposide (+) weight (-) Male (-) TNM-N0 (+) TNM-M1a (+)	TNM-N2 (-)		ECOG 2 (+) TNM-N0 (-) LDH (+)

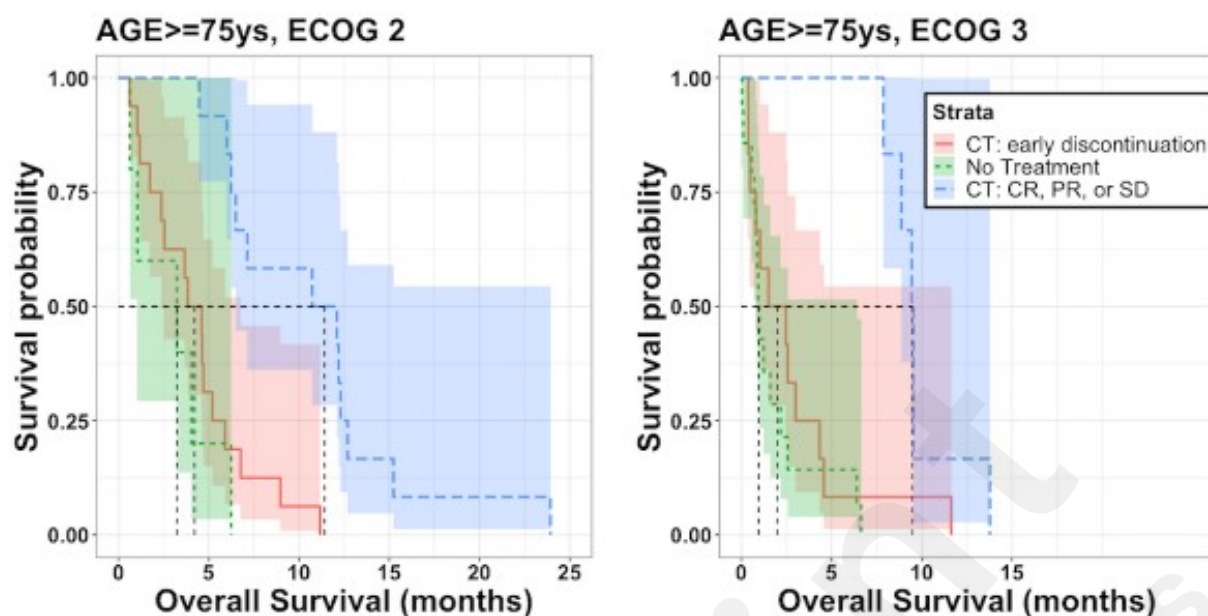


Figure 4. Kaplan Meier curves of patients older than 75 years (yrs) with ECOG performance status (PS) 2-3 stratified by good response to first line chemotherapy, early discontinuation of treatment, and patients that did not received chemotherapy. CT: chemotherapy, CR: complete response, PR: partial response, SD: stable disease.

Discussion

In this study we analysed a longitudinal real-world cohort of SCLC patients and their treatment pathways. The analysis allowed the study of the relationships between treatment decisions, phases of progression of disease, occurrence of AEs, and overall survival. To the best of our knowledge, this is one of the most detailed longitudinal datasets of SCLC real-world treatment decisions. This type of data aims to add novel insights on dosing effects over time, emergence of toxicity, and the decision points relative to the clinical outcomes.

The improved prognosis of patients that experienced neutropenia align with previous findings^{41,42}. The analysis allowed the investigation of treatment-related adverse events in relation to efficacy of therapy or detect when this was instead linked to deterioration of the patient conditions. For example, the cumulative incidence curves of AEs stratified by responding patients against early discontinuation of the therapy in Figure 2 showed the role of neutropenia when the time windows of AEs may overlap with the efficacy of treatment (Figure 3).

We identified significant ORs relating to the dose of etoposide and increased occurrence of AEs. The analysis showed a similar OS for patients receiving four doses during first line therapy, regardless of whether the patient had dose adjustment/reduction or not (Figure 3). Worse prognosis was observed when there was a deterioration in ECOG PS or early discontinuation of treatment.

We interpret our findings as indicating that the occurrence of neutropenia reflects a longer duration of treatment exposure, rather than being a causal factor for improved survival. Patients who completed all four treatment cycles—often at reduced doses—demonstrated better outcomes than those receiving fewer cycles at higher doses. This is supported by the higher incidence of neutropenia among patients with treatment response, suggesting that extended exposure, not

intensity, is the key contributor to therapeutic benefit.

SCLC treatment effects vary widely between individuals. This could suggest the to identify key covariates and clinically relevant subgroups that might inform a more tailored dosing , or in some cases, guide decisions against treatment altogether, in order to improve survival while avoiding AEs in patients unlikely to benefit from therapy^{10,11,43} . Indeed, etoposide dose calculation is based on BSA that has been shown to present such limitations⁴⁴ . Similar analyses could be done for the second line. In fact, only a few patients received the full expected dose without dose reduction, and the platinum dose and PE treatment were strongly associated with AEs.

ECOG PS 2-3 and elevated LDH serum levels are known to be prognostic factors of worse OS in SCLC^{8,45} . Our results, obtained by the careful analysis of the longitudinal data line by line with the clinical experts, added a new insight by showing their role in predicting early discontinuation of treatment, which was associated with significantly poorer OS.

Another finding was the high probability of a favourable response to first-line therapy in male patients, along with their lower likelihood of experiencing AEs (Table 3-4 and Supplementary Tables S4-5). Moreover, when stratifying older patients by sex, women who developed neutropenia exhibited survival rates comparable to those in the same subgroup who did not receive chemotherapy. A possible explanation may be that females patients is a group that historically has been underrepresented in clinical trials^{9,46,47} . In addition, in the second-line setting, the worse PFS and high frequent early discontinuation for male patients might suggest the need of further studies regarding different therapy responses for males and females.

The high-risk subgroup analysis allowed detection of key groups of patients such as elderly patients with poor ECOG PS, where an eventual poor response to therapy is associated with similar survival as compared to patients not receiving chemotherapy. This finding is clinically relevant, as it highlights the importance of balancing benefit and risk of treating such fragile patients and may help clinicians avoid overtreating this vulnerable population. These outcomes underline the potential role that real-world data may fill in providing evidence on treatment in groups of patients that are underrepresented in clinical trials⁹ .

The design of the analytic pipeline relied on the consideration that the data should be analysed using models capable of capturing the complexity of the patient treatment pathways without sacrificing interpretability (such as, black box machine learning models²⁴). We believe that this is an important consideration when approaching real-world data. In fact, the discussions with the clinicians on the obtained results allowed the assessment of the validity of the results, the detection of potential inherent bias, and confounders related to operations and non-observable effects not captured from the data. For example, it was observed that the 10-day dosing regimen of etoposide resulted in poorer survival. At the time of the present analysis this dosing regimen had however already been abandoned due to toxicity caused by the limited time for patients to recover between the cycles (Tables 2-3).

The study had several limitations. The reason for missing values was not at random for some of the variables, thus introducing some biases on the results. Particularly for baseline neutrophil counts, where missing observations were high across all treatment lines.

Further, these variables were available only at the start of the treatment line, meaning that we were not able to analyse their time-profiles (i.e., neutrophil

counts over time between doses). Moreover, it should be stated that AEs are not always described in medical notes, thus introducing investigator bias in the collected data.

Another limitation was the sample size of the high-risk subgroups and patients receiving the second line treatment, resulting in higher estimates of variability of the outcomes as compared to the first line treatment cohort. In detail, the odds ratios of the mixed effects models resulted in wide confidence intervals. The reason of this high variability in outcomes could be since these small cohorts were also associated with fewer longitudinal observations per patient (the number of administrated doses). Moreover, patients receiving more than two lines were not included in the multivariate analysis as this was only observed in few instances.

The median age of our population was 72 years, yet we chose 75 years to better isolate a more vulnerable subset of patients by moving further from the median. This aligns with existing studies that identify this age group as particularly fragile, often requiring specialized considerations in treatment planning⁴⁸. Future studies exploring different criteria and strategies to separate elderly patients would be required.

This was a single-centre study and expanding the analysis by including other SCLC cohorts from other centres would be beneficial. A comparison of the dose adaptation and the follow-up processes in clinical trials would be of high importance⁹. A future aspect to explore would be to expand the collection of longitudinal real-world data, designed similarly to the study presented here, in larger initiatives such as multi-centre studies or cancer registries⁴⁹⁻⁵¹, and recent new proposed therapies (i.e., Durvalumab after chemoradiotherapy⁵², Tarlatamab/Lubri in second and third line settings⁵³, and immune-oncology therapies⁸).

This study could inform the design of future works, such as multi-state models of patient survival in function of different therapies and events^{54,55}, or model-informed precision dosing models to explore individualised treatment and dosing to reduce the risk of AEs (e.g., neutropenia) without compromising the efficacy of the treatment in terms of survival outcomes^{56,57}. Furthermore, the outcomes from the approaches presented herein can inform more reliable future analyses using for example machine learning algorithms suitable for modelling longitudinal processes, such as physics-based neural networks⁵⁸.

A continued fruitful collaboration between clinical and technical experts when analysing complex data and contextualising the clinical processes, as we did in this work, would enhance the impact of real-world data in clinical practice and clinical drug development.

Author contributions

L.M. and A.D. wrote the manuscript. L. M., A.S.D., A.D., L.D.P., S.T., J.R., R.L., and S.M. designed the research. L.M., A. D., L.D.P. and A.S.D. performed the research. L.M., A.D., A.S.D., and L.D.P. analysed the data.

Competing interests

The authors declare no competing interests.

Funding information

The Swedish Cancer Society (grant no. CAN 2018/597 and CAN2021/1469 Pj01) to R. Lewensohn and from the Stockholm Cancer Society (grant no. #201202 to R. Lewensohn and #174063, #201103 and #231123 to L. De Petris).

Data availability

The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the authors upon reasonable request. Data are located in a controlled access data storage at Karolinska University Hospital (contact: luigi.de.petris@ki.se).

References

1. Pietanza MC, Byers LA, Minna JD, Rudin CM. Small Cell Lung Cancer: Will Recent Progress Lead to Improved Outcomes? *Clinical Cancer Research*. 2015;21(10):2244-2255. doi:10.1158/1078-0432.CCR-14-2958
2. Reddy HG, Qin A, Kalemkerian GP. Emerging drugs for small cell lung cancer: a focused review on immune checkpoint inhibitors. *Expert opinion on emerging drugs*. 2020;25(3):353-366. doi:10.1080/14728214.2020.1798929
3. Kalemkerian GP, Loo BW, Akerley W, et al. NCCN Guidelines Insights: Small Cell Lung Cancer, Version 2.2018. *Journal of the National Comprehensive Cancer Network*. 2018;16(10):1171-1182. doi:10.6004/jnccn.2018.0079
4. Dingemans AMC, Früh M, Ardizzoni A, et al. Small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up☆. *Annals of Oncology*. 2021;32(7):839-853. doi:10.1016/j.annonc.2021.03.207
5. Horn L, Mansfield AS, Szczenińska A, et al. First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *New England Journal of Medicine*. 2018;379(23):2220-2229. doi:10.1056/NEJMoa1809064
6. Tendler S, Zhan Y, Pettersson A, et al. Treatment patterns and survival outcomes for small-cell lung cancer patients - a Swedish single center cohort study. *Acta Oncol*. 2020;59(4):388-394.

doi:10.1080/0284186X.2019.1711165

7. Ardizzoni A, Tiseo M, Boni L. Validation of standard definition of sensitive versus refractory relapsed small cell lung cancer: A pooled analysis of topotecan second-line trials. *European Journal of Cancer*. 2014;50(13):2211-2218. doi:10.1016/j.ejca.2014.06.002
8. Johal S, Hettle R, Carroll J, Maguire P, Wynne T. Real-world treatment patterns and outcomes in small-cell lung cancer: a systematic literature review. *J Thorac Dis*. 2021;13(6):3692-3707. doi:10.21037/jtd-20-3034
9. Marzano L, Darwich AS, Dan A, et al. Exploring the discrepancies between clinical trials and real-world data: A small-cell lung cancer study. *Clinical and Translational Science*. 2024;17(8):e13909. doi:10.1111/cts.13909
10. Darwich AS, Ogungbenro K, Vinks AA, et al. Why Has Model-Informed Precision Dosing Not Yet Become Common Clinical Reality? Lessons From the Past and a Roadmap for the Future. *Clinical Pharmacology & Therapeutics*. 2017;101(5):646-656. doi:10.1002/cpt.659
11. Darwich AS, Polasek TM, Aronson JK, et al. Model-Informed Precision Dosing: Background, Requirements, Validation, Implementation, and Forward Trajectory of Individualizing Drug Therapy. *Annual Review of Pharmacology and Toxicology*. 2021;61(Volume 61, 2021):225-245. doi:10.1146/annurev-pharmtox-033020-113257
12. Povsic M, Enstone A, Wyn R, Kornalska K, Penrod JR, Yuan Y. Real-world effectiveness and tolerability of small-cell lung cancer (SCLC) treatments: A systematic literature review (SLR). *PLoS ONE*. 2019;14(7):e0219622. doi:10.1371/journal.pone.0219622
13. Moser SS, Bar J, Kan I, et al. Real World Analysis of Small Cell Lung Cancer Patients: Prognostic Factors and Treatment Outcomes. *Curr Oncol*. 2021;28(1):317-331. doi:10.3390/curroncol28010036
14. Zellweger NM, Schmid S, Bertschinger M, et al. 1540P Real-world analysis of outcomes of first-line chemo-immunotherapy in patients with extensive disease small cell lung cancer (ED-SCLC). *Annals of Oncology*. 2022;33:S1251. doi:10.1016/j.annonc.2022.07.1634
15. Derksen JWG, May AM, Koopman M. The era of alternative designs to connect randomized clinical trials and real-world data. *Nat Rev Clin Oncol*. 2019;16(9):589-589. doi:10.1038/s41571-019-0250-0
16. Baumfeld Andre E, Carrington N, Siami FS, et al. The Current Landscape and Emerging Applications for Real-World Data in Diagnostics and Clinical Decision Support and its Impact on Regulatory Decision Making. *Clinical Pharmacology & Therapeutics*. 2022;112(6):1172-1182. doi:10.1002/CPT.2565
17. Schurman B. The Framework for FDA's Real-World Evidence Program. *Applied Clinical Trials*. 2019;28(4):15-17.
18. Hiramatsu K, Barrett A, Miyata Y. Current Status, Challenges, and Future Perspectives of Real-World Data and Real-World Evidence in Japan. *Drugs - Real World Outcomes*. 2021;8(4):459-480. doi:10.1007/S40801-021-00266-3/TABLES/8

19. Dang A. Real-World Evidence: A Primer. *Pharmaceutical Medicine*. 2023;37(1):25-36. doi:10.1007/S40290-022-00456-6/TABLES/3
20. Greenhalgh T, Papoutsis C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. *BMC Medicine*. 2018;16(1):95. doi:10.1186/s12916-018-1089-4
21. Thatcher N, Girling DJ, Hopwood P, et al. Improving Survival Without Reducing Quality of Life in Small-Cell Lung Cancer Patients by Increasing the Dose-Intensity of Chemotherapy With Granulocyte Colony-Stimulating Factor Support: Results of a British Medical Research Council Multicenter Randomized Trial. *JCO*. 2000;18(2):395-395. doi:10.1200/JCO.2000.18.2.395
22. Socinski MA, Smit EF, Lorigan P, et al. Phase III study of pemetrexed plus carboplatin compared with etoposide plus carboplatin in chemotherapy-naïve patients with extensive-stage small-cell lung cancer. *Journal of Clinical Oncology*. 2009;27(28):4787-4792. doi:10.1200/JCO.2009.23.1548
23. Ibrahim JG, Chu H, Chen LM. Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *JCO*. 2010;28(16):2796-2801. doi:10.1200/JCO.2009.25.0654
24. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell*. 2019;1(5):206-215. doi:10.1038/s42256-019-0048-x
25. Alabi RO, Youssef O, Pirinen M, et al. Machine learning in oral squamous cell carcinoma: Current status, clinical concerns and prospects for future—A systematic review. *Artificial Intelligence in Medicine*. 2021;115:102060. doi:10.1016/j.artmed.2021.102060
26. Aubel P, Antigny M, Fougeray R, Dubois F, Saint-Hilary G. A Bayesian approach for event predictions in clinical trials with time-to-event outcomes. *Statistics in Medicine*. 2021;40(28):6344-6359. doi:10.1002/sim.9186
27. Yin J, Qin R, Ezzalfani M, Sargent DJ, Mandrekar SJ. A Bayesian dose-finding design incorporating toxicity data from multiple treatment cycles. *Statistics in Medicine*. 2017;36(1):67-80. doi:10.1002/sim.7134
28. Zhao D, Zhu J, Wang L. Bayesian interval-based oncology dose-finding design with repeated quasi-continuous toxicity model. *Contemporary Clinical Trials*. 2021;102:106265. doi:10.1016/j.cct.2021.106265
29. Xu R. Measuring explained variation in linear mixed effects models. *Statistics in Medicine*. 2003;22(22):3527-3541. doi:10.1002/sim.1572
30. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982;38(4):963-974. doi:10.2307/2529876
31. Coz E, Fauvernier M, Maucourt-Boulch D. An Overview of Regression Models for Adverse Events Analysis. *Drug Saf*. 2024;47(3):205-216. doi:10.1007/s40264-023-01380-7
32. Dignam JJ, Zhang Q, Kocherginsky M. The Use and Interpretation of Competing Risks Regression Models. *Clinical Cancer Research*. 2012;18(8):2301-2308. doi:10.1158/1078-0432.CCR-11-2097

33. Calvert AH, Newell DR, Gumbrell LA, et al. Carboplatin dosage: prospective evaluation of a simple formula based on renal function. *J Clin Oncol.* 1989;7(11):1748-1756. doi:10.1200/JCO.1989.7.11.1748
34. Jiang S, Huang L, Zhen H, Jin P, Wang J, Hu Z. Carboplatin versus cisplatin in combination with etoposide in the first-line treatment of small cell lung cancer: a pooled analysis. *BMC Cancer.* 2021;21(1):1308. doi:10.1186/s12885-021-09034-6
35. Harrell FE. Cox Proportional Hazards Regression Model. In: *Regression Modeling Strategies*. Springer Series in Statistics. Springer International Publishing; 2015:475-519. doi:10.1007/978-3-319-19425-7_20
36. Suchmacher M, Geller M. Chapter 2 - Disease frequency measures. In: Suchmacher M, Geller M, eds. *Practical Biostatistics (Second Edition)*. Academic Press; 2021:9-15. doi:10.1016/B978-0-323-90102-4.00017-5
37. Bayesian Multivariate Logistic Regression | Biometrics | Oxford Academic. Accessed August 16, 2024. <https://academic.oup.com/biometrics/article/60/3/739/7289337>
38. Mixed-Effects Regression Model for Longitudinal Multivariate Ordinal Data | Biometrics | Oxford Academic. Accessed August 16, 2024. <https://academic.oup.com/biometrics/article/62/1/261/7306426>
39. Tang F, Ishwaran H. Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal.* 2017;10(6):363-377. doi:10.1002/sam.11348
40. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of Random Forest and Parametric Imputation Models for Imputing Missing Data Using MICE: A CALIBER Study. *American Journal of Epidemiology.* 2014;179(6):764-774. doi:10.1093/aje/kwt312
41. Kasi PM, Grothey A. Chemotherapy-Induced Neutropenia as a Prognostic and Predictive Marker of Outcomes in Solid-Tumor Patients. *Drugs.* 2018;78(7):737-745. doi:10.1007/s40265-018-0909-3
42. Miyagi T, Tsuji D, Kawasaki Y, et al. Chemotherapy-induced neutropenia as a prognostic factor in patients with extensive-stage small cell lung cancer. *Eur J Clin Pharmacol.* 2023;79(3):407-414. doi:10.1007/s00228-023-03451-1
43. Sachs JR, Mayawala K, Gadamsetty S, Kang SP, de Alwis DP. Optimal Dosing for Targeted Therapies in Oncology: Drug Development Cases Leading by Example. *Clinical Cancer Research.* 2016;22(6):1318-1324. doi:10.1158/1078-0432.CCR-15-1295
44. Sawyer M, Ratain MJ. Body Surface Area as a Determinant of Pharmacokinetics and Drug Dosing. *Invest New Drugs.* 2001;19(2):171-177. doi:10.1023/A:1010639201787
45. Zhang X, Guo M, Fan J, et al. Prognostic significance of serum LDH in small cell lung cancer: A systematic review with meta-analysis. *Cancer Biomark.* 2016;16(3):415-423. doi:10.3233/CBM-160580
46. Singh S, Parulekar W, Murray N, et al. Influence of Sex on Toxicity and Treatment Outcome in Small-Cell Lung Cancer. *JCO.* 2005;23(4):850-856. doi:10.1200/JCO.2005.03.171

47. Pang HH, Wang X, Stinchcombe TE, et al. Enrollment Trends and Disparity Among Patients With Lung Cancer in National Clinical Trials, 1990 to 2012. *JCO*. 2016;34(33):3992-3999. doi:10.1200/JCO.2016.67.7088
48. Surbone A, Kagawa-Singer M, Terret C, Baider L. The illness trajectory of elderly cancer patients across cultures: SIOG position paper. *Annals of Oncology*. 2007;18(4):633-638. doi:10.1093/annonc/mdl178
49. Wormeli P, Mazreku J, Pine J, Damesyn M. Next Generation of Central Cancer Registries. *JCO Clin Cancer Inform*. 2021;(5):288-294. doi:10.1200/CCI.20.00177
50. Zanetti R, Schmidtman I, Sacchetto L, et al. Completeness and timeliness: Cancer registries could/should improve their performance. *European Journal of Cancer*. 2015;51(9):1091-1098. doi:10.1016/j.ejca.2013.11.040
51. Leon L, Kosty M, Jahanzeb M, et al. Effectiveness of bevacizumab exposure beyond disease progression in patients with non-small-cell lung cancer: analyses of the ARIES observational cohort study. *Pharmacoeconomics and Drug Safety*. 2016;25(5):569-577. doi:10.1002/pds.3948
52. Cheng Y, Spigel DR, Cho BC, et al. Durvalumab after Chemoradiotherapy in Limited-Stage Small-Cell Lung Cancer. *New England Journal of Medicine*. 2024;391(14):1313-1327. doi:10.1056/NEJMoa2404873
53. Ahn MJ, Cho BC, Felip E, et al. Tarlatamab for Patients with Previously Treated Small-Cell Lung Cancer. *New England Journal of Medicine*. 2023;389(22):2063-2075. doi:10.1056/NEJMoa2307980
54. Krishnan SM, Friberg LE, Mercier F, et al. Multistate Pharmacometric Model to Define the Impact of Second-Line Immunotherapies on the Survival Outcome of the IMpower131 Study. *Clinical Pharmacology & Therapeutics*. Published online January 29, 2023. doi:10.1002/CPT.2838
55. Krishnan SM, Friberg LE, Bruno R, Beyer U, Jin JY, Karlsson MO. Multistate model for pharmacometric analyses of overall survival in HER2-negative breast cancer patients treated with docetaxel. *CPT: Pharmacometrics & Systems Pharmacology*. 2021;10(10):1255-1266. doi:10.1002/psp4.12693
56. Lombard A, Mistry H, Aarons L, Ogungbenro K. Dose individualisation in oncology using chemotherapy-induced neutropenia: Example of docetaxel in non-small cell lung cancer patients. *British Journal of Clinical Pharmacology*. 2021;87(4):2053-2063. doi:10.1111/bcp.14614
57. Liu H, Ibrahim EIK, Centanni M, Sarr C, Venkatakrishnan K, Friberg LE. Integrated modeling of biomarkers, survival and safety in clinical oncology drug development. *Advanced Drug Delivery Reviews*. 2025;216:115476. doi:10.1016/j.addr.2024.115476
58. Laurie M, Lu J. Explainable deep learning for tumor dynamic modeling and overall survival prediction using Neural-ODE. *npj Syst Biol Appl*. 2023;9(1):1-11. doi:10.1038/s41540-023-00317-1



Supplementary Files

Multimedia Appendixes

Supplementary Materials part 1 (Figures).

URL: <http://asset.jmir.pub/assets/0dc2f5494c466d490b39c0757aa330e1.docx>

Supplementary Tables.

URL: <http://asset.jmir.pub/assets/feb87950eb9fc71794c3ad25090c39da.xlsx>