PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Evaluating transportability of overall survival estimates from US to UK populations receiving first-line treatment for advanced non-small cell lung cancer: a retrospective cohort study

Authors

Kent, Seamus; Mpofu, Philani; Duffield, Stephen; Adam, Jane; Beal, Brennan; Royce, Trevor J.; Adamson, Blythe J. S.; Kasturi, Jyotsna; Sujenthiran, Arun; Jonsson, Pall

---

VERSION 1 - REVIEW

Reviewer	**1**
Name	**Offorha, Bright Chiemezie**
Affiliation Research	**The University of Sheffield, School of Health and Related**
Date	**21-Aug-2024**
COI	**None.**

This study was robustly conducted to circumvent the unique challenges of observational studies. All aspects of the research are well-reported, making it easy to understand. The authors put a great effort into this project and it is commendable.

I have a few suggestions, could the authors further describe the matching method used - matching adjusted indirect comparison?

Could the authors use different line types in the survival curves, so that they can be distinguishable when printed in black and white?

VERSION 1 - AUTHOR RESPONSE

Reviewer: 1
Dr. Bright Chiemezie Offorha, The University of Sheffield
Comments to the Author:

This study was robustly conducted to circumvent the unique challenges of observational studies. All aspects of the research are well-reported, making it easy to understand. The authors put a great effort into this project and it is commendable.

I have a few suggestions, could the authors further describe the matching method used - matching adjusted indirect comparison?

Thank you for the helpful suggestion to explain the methods more clearly. We used the matching-adjusted indirect comparison (MAIC), which is common when researchers have access to patient-level data in one country and aggregate population-level data in a different country. We have now revised the manuscript to explain this method more clearly.

Edits in the document (lines: 174 to 182)
"We used the matching-adjustment indirect comparison (MAIC) approach to standardise the characteristics of US patients to those of UK patients represented in Lester et al. 2021. We selected MAIC because it enabled us to standardise individual patient data from the US using summary/published data from the UK. MAIC estimated weights to ensure that the average characteristics of the US study population matched those of the UK study population."

Could the authors use different line types in the survival curves, so that they can be distinguishable when printed in black and white?

Author response:

Thank you for the suggestion. We have updated the plots with different colour palettes and linetypes. The new versions should be clear whether printed in grayscale or colour.

Reviewer: 2
Dr. Ronald Damhuis, Integraal Kankercentrum Nederland
Comments to the Author:

2024-085722
Interesting exercise to compare OS after 1L SACT for advanced NSCLC between the UK and the US. Two tables (1 filling 3 pages), one figure (1a/d), 9 supplementary tables, 2 supplementary figures. Methods, results and intentions are well-described.

Thank you for the encouraging comment. We agree this was a really interesting exercise to do and we have learned a lot that we are happy to share with readers.

UK data are based on (Lester,2021) at a time when immuno mono was reimbursed for PD-L1>50. In table 1, PD-L1 information is mainly missing. Is there any information

about the use of the PD-L1>50 cutoff in the USA? Can we assume that the results in the USA also reflect a PD-L1>50 population?

It is very insightful to notice how the label indications for immunotherapy were slightly different between the two countries during the time period of this study. We added this detail to the manuscript so readers understand that the main analysis compares all US patients regardless of cutoff threshold. Based on your comment we have added a new sensitivity analysis to the paper and provide a more detailed explanation below.

PDL1 Missingness

In the current manuscript, the PD-L1 results for the U.S. cohort are based on the PD-L1 interpretations recorded in the medical charts, hence the large missingness for the variable. We have now updated the PD-L1 results in the US cohort to reflect a composite of PD-L1 tumour proportion score and the PD-L1 biomarker status as assigned in medical charts. Under this composite definition, one is considered to be PD-L1 positive if there is reference to PD-L1 positivity in the medical chart or if PD-L1 staining result has a TPS of ≥1%.

Immunotherapy-exposed cohort with PDL1 (≥50%)

Of the 836 patients in the US immunotherapy-exposed cohort, 603 (72%) had a PD-L1 tumour proportion score (TPS) of ≥50%. The baseline characteristics of the PDL1 50+ cohort were largely similar to those of the whole immunotherapy-exposed cohort that was included in the main analysis. We used MAIC to standardise the US PDL1 50+ cohort to the UK cohort. The variables used in standardisation were the same as those used in the main analysis. The comparison of OS between the UK cohort and the US PDL1 50+ cohort (before and after standardisation) is presented in the table below.

Analysis	Summary	US PDL1 50+ unweighted	US PDL1 50+ weighted	UK
IO mono.	mOS (95% CI)	11.6(10.0–14.9)	14.9 (11.7–18.9)	14.0 (10.7–20.6)
	12 months RMST (se)	8.01 (0.19)	8.48 (0.19)	8.79 (0.31)
	24 months RMST (se)	13.12 (0.40)	14.03 (0.42)	14.23 (0.69)

US data standardised to reflect average characteristics of patients in the UK for age, sex, ECOG PS score (0–1 or 2+), and histology (squamous cell, non-squamous cell, unknown). CI=confidence interval. IO mono=immunotherapy monotherapy. mOS=median overall survival. RMST=restricted mean survival time. se=standard error.

The survival estimates for the PDL1 50+ cohort were slightly higher than those of the whole immunotherapy-exposed cohort that was considered for the main analysis.

Based on the above supplementary sensitivity analysis, we have added the following inline edit to the manuscript (Lines 338 to 341):

"During the time of the study, there were differences between the countries in biomarker testing threshold scores for use of immunotherapy, though additional sensitivity analysis did not find this to meaningfully change the study results (see Supplementary Table 10)."

I was surprised by the major difference in the proportion of ECOG>=2 for 1L IO between US and UK. For chemotherapy, ECOG>=2 proportions seem similar. ECOG is sensitive to inter-observer variation. Can you reflect on the importance of variation in ECOG-definitions when comparing US and UK data, especially when ECOG may be used as a parameter in NICE recommendations?

Thank you for your insightful observation. The higher-than-expected frequency of patients with ECOG scores of 0-1 in the UK study is indeed noteworthy. This may be attributed to the timing of the study, which was conducted shortly after the introduction of immunotherapies in the UK. During this initial phase, it is likely that the patient population consisted mainly of healthier individuals. As these treatments become more widely accessible and established over time, we anticipate an increase in the proportion of patients with an ECOG score of 2 or higher. It could also reflect country-level preferences of patients and healthcare providers, with a greater tendency to offer systemic therapy to patients with a poorer performance status in the US compared to the UK.

TTD is only described for the US population, apparently due to 'difference in definitions'. This explanation is not clear to me. For oral drugs, this may be an issue, but for systemic treatment this may (merely) differ by 21 days. Lasala compared TTD studies and states that most studies do not even mention the definition https://www.tandfonline.com/doi/full/10.1080/03007995.2023.2192610. Please clarify why comparison was not feasible.

Below we provide more detail to explain how the definitions of real-world time to treatment discontinuation (rwTTD) are different. This table has been added to the supplementary appendix (Supplementary Table 11) of the paper (cited in line 347). A key difference in data curation models is that Flatiron curates treatment episodes rather than cycles, which are a different measurement unit of time on treatment. We will try to do this in future studies where we have access to patient-level data in both countries as it will be helpful to answer.

	Definition of Endpoint in data source	
Endpoint	**Flatiron Health**	**Lester et al. 2021**
rwTTD	Time from 1L treatment initiation to treatment discontinuation (for any reason including death).	Time from 1L treatment initiation to treatment discontinuation (for any reason including death).
	Start date: first drug episode for the drug of interest within a given line of therapy (LOT) **End date:** last drug episode for the drug of interest within a given LOT Time at risk is time elapsed between start and end dates of a LOT	["...in patients who discontinued treatment but were still alive, the treatment *end date was recorded as the start date of the last treatment cycle* because a definitive end date of the last cycle was not available, and the last cycle start date was the latest date when it was certain that treatment was continuing."]
	A patient is treated as uncensored if ANY of the following three events are observed in the data: The patient advanced to a new LOT. Because rwTTD is defined within a given LOT, evidence of advancement to a new LOT mandates the discontinuation of the treatment offered under the preceding line. The patient has not advanced to a new LOT, but has a recorded date of death. Mortality should be treated as confirmatory of treatment discontinuation. The patient has not advanced to a new LOT and has no recorded date of death, but has sufficient evidence of confirmed structured activity after the last drug episode for the drug(s) of interest. In the absence of a more definitive condition like LOT advancement or evidence of death, inference of treatment discontinuation from structured EHR data is necessary. A prolonged period (e.g., 120 days) of confirmed structured activity following the last recorded drug episode may be considered reasonable evidence of treatment discontinuation because it suggests that the patient is still being followed at the treating clinic; thus, one can assume consistent capture of treatment data. As such, it is unlikely that the cessation of new drug episodes is the result of missing data.	

Table 2. After targeted in column2, RMST (se) is not mentioned.

This was indeed an accidental omission; thank you for pointing this out to us. We have added "RSMT (se)" to the targeted therapy results in Table 2, so the results description is now complete.

A similar phenomena (is this plural or singular?)

You are right. The correct word should be "phenomenon" given that we are making a singular reference. We have corrected the grammatical error (line 311).

VERSION 2 - REVIEW

Reviewer	**2**
Name	**Damhuis, Ronald**
Affiliation	**Integraal Kankercentrum Nederland**
Date	**12-Nov-2024**
COI	

Interesting paper for gourmets