Identification of treatment intent from the actual time-to-treatment distribution in prostate cancer patients

Asieh Golozar[1,2], Christian Reich[2,3]
[1]Regeneron Pharmaceuticals NY USA, [2]OHDSI Oncology WG, [3]IQVIA MA USA

Background

After diagnosis, prostate cancer patients face a crucial decision whether to undergo treatment to combat the disease right away, or whether to opt for conservative management of the disease and delay treatment. This is based on the recognition that death from other causes might occur before a malignancy of little activity shows symptoms, especially in older multimorbid patients, where sparing the treatment might avoid an enormous burden on the patient. There are two strategies for delaying treatment: Watchful Waiting (WW) is the option for patients unsuitable for curative treatment and in no urgent need for palliative treatment until the occurrence of symptoms. Active Surveillance (AS) on the other hand involves regular monitoring of disease progression and initiating curative therapy when the cancer progresses.

It is an important task of clinical research to inform this decision at the time of diagnosis. For that, patients with the different choices must be followed-up. Since deferral could take as long as several years, retrospective studies using large observational data, such as those from claims and electronic health records (EHR), could play a vital role in evidence generation.

However, the decision about immediate treatment or deferral according to WW or AS is rarely captured in the data, and patients therefore cannot be assigned to study cohorts. One solution is based on the actual time between diagnosis and treatment. The challenge of this approach is that there is no obvious or generally accepted cut-off time after which a treatment can be designated "deferred".

To address the first problem, a data driven approach might help distinguish between patients with the two choices. The time to treatment (TTT) of the "immediate" treatment population depends how speedy the diagnostic workup can be completed and treatment initiated by the provider. The TTT of the "deferred" population stretches out while no sign of cancer activity is observed. Therefore, the actual TTT distribution should be composed of two overlapping distributions with different parameters.

In this study, we attempted to empirically identify the two distinct populations from the data and to determine the optimal cut-off, minimizing misclassification of the patients and potential selection bias.

Methods

Five large scale observational databases were used containing patients newly diagnosed with prostate cancer: IQVIA PharMetrics Plus, Open Claims, Hospital Charge Data Master, Ambulatory EMR and Oncology EMR. We performed curve fitting to finite mixture models

including Weibull, Gamma and Log-normal via expectation maximization (EM) and selected the best model for each database using the Bayesian information criterion (BIC). The parameters of the selected models were estimated using maximum likelihood estimation (MLE).

Results

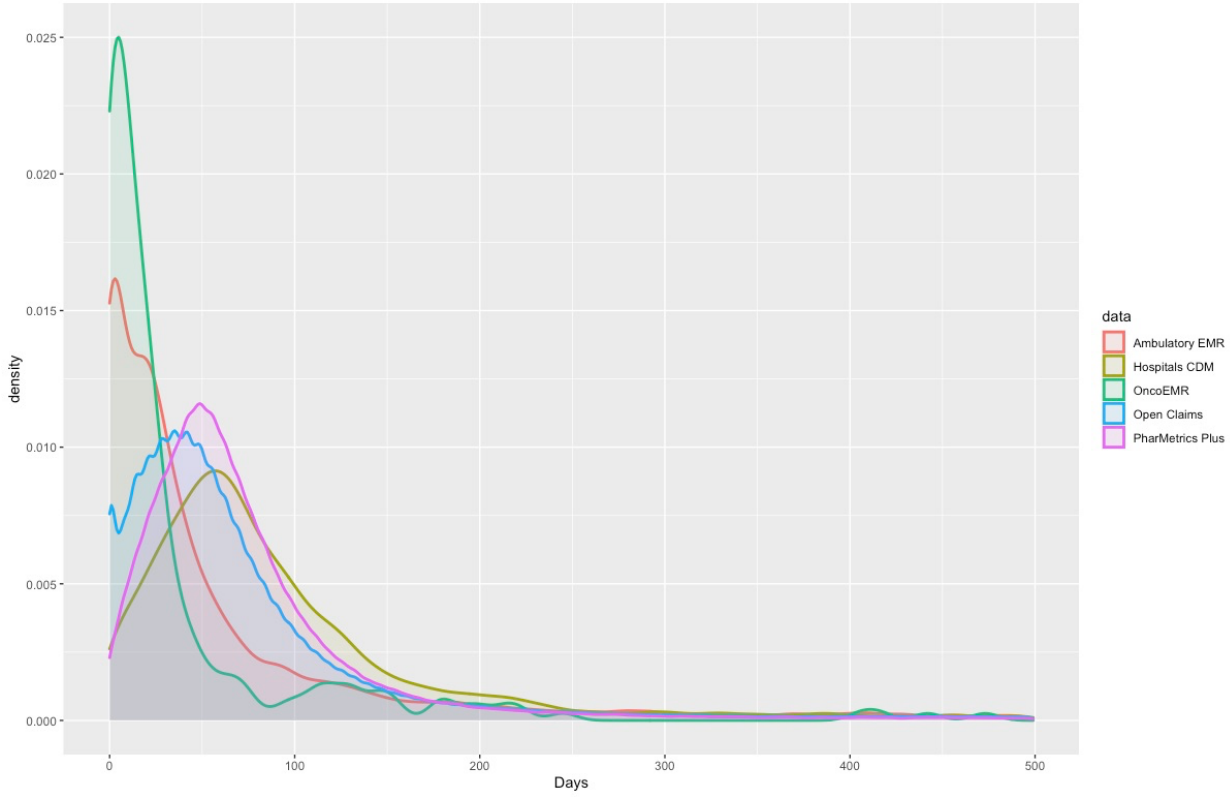
A total of 912,789 newly diagnosed prostate cancer patients across a network of claims and EHR data were included in the study (Table 1).

Table 1. Time to treatment initiation and follow-up in the participating databases

	N. Patients	Median Time to Treatment initiation, days (IQR)	Median follow-up time, days (IQR)
IQVIA Ambulatory EMR	9791	31 (30, 33)	980 (957, 1001)
IQVIA Hospital CDM	7476	73 (71, 75)	753 (727, 773)
IQVIA Oncology EMR	219	14 (10, 18)	359 (295, 457)
IQVIA Open Claims	692286	55 (55, 55)	1899 (1895, 1904)
IQVIA PharMetrics Plus	203017	61 (60, 61)	950 (944, 955)

Time to treatment initiation was variable across different databases and the overall irregular shape seems to indicate a multi-modal distribution (Fig 1).

Fig 1. Time to Initiation of Treatment Following Initial Diagnosis of Prostate Cancer



An empirical test of that assessment through fitting of the model showed the superiority of the bimodal distribution, supporting the hypothesis that the data contain two distinct populations (Fig 2, only best fitting Weibull model shown).

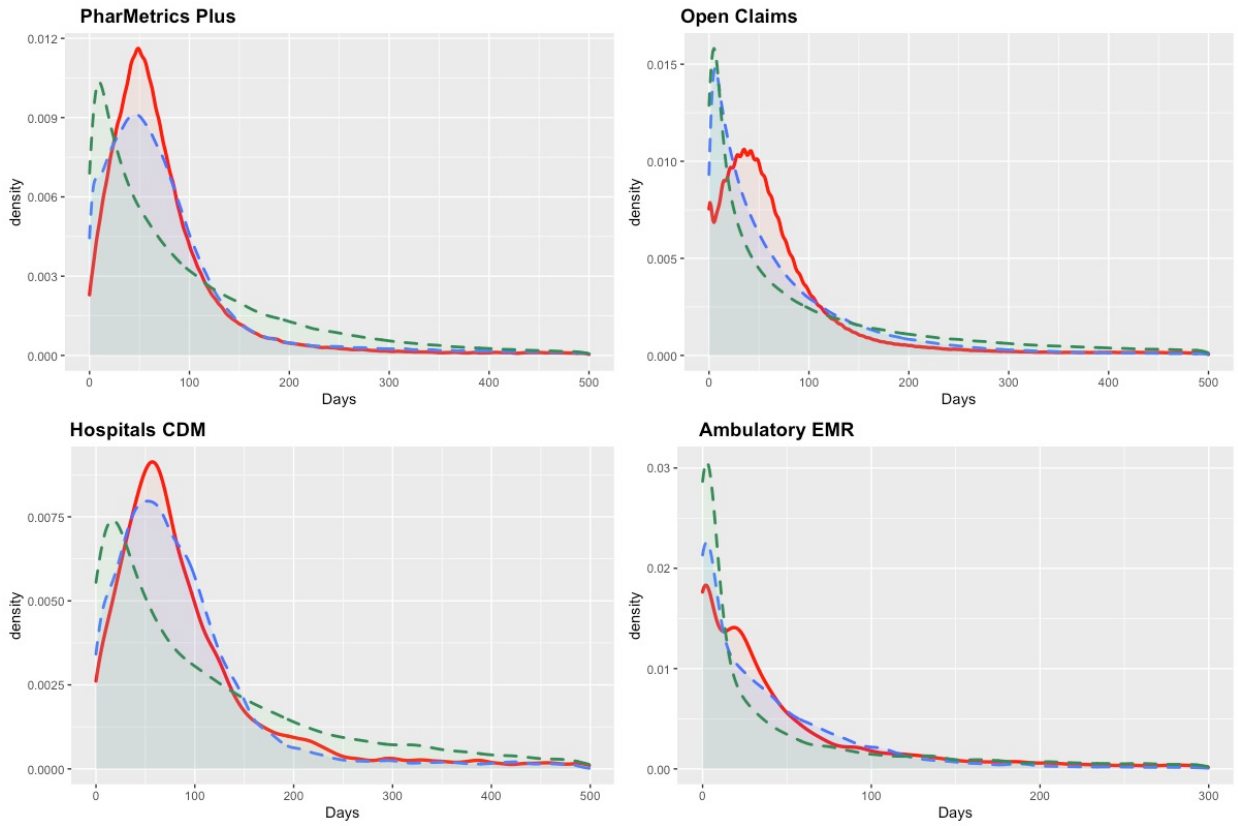


Fig 2. Fitted unimodal (green) and bimodal (blue) Weibull distributions together with the observed data (red).

From the selection of the parametric distributions assessed in this work, a bimodal two-parameter Weibull distribution was the best fit to the data. in all five databases. The estimated shape and scales varied across databases and are summarized in Table 2.

Table 2. EM estimated parameters for the TTT distribution.

		Component 1	Component 2
PharMetrics Plus	**Proportions**	0.8	0.2
	Mean	63.4	295.6
	Shape	1.9	0.6
	Scale	71.4	189.2
Open Claims	**Proportions**	0.9	0.1
	Mean	67.8	1067.5
	Shape	0.9	1.1
	Scale	64.1	1111.3
Ambulatory EMR	**Proportions**	0.5	0.5
	Mean	1025.1	49.4
	Shape	0.2	1.3
	Scale	29.3	53.2
Hospital CDM	**Proportions**	0.7	0.3
	Mean	76.6	336.4
	Shape	1.9	0.6
	Scale	86.4	199.6

We then assessed if the two populations depicted from the two components are distinct enough to allow the definition of cohorts using a single TTT cut-off (Fig 3). We found that (i) the distribution of the two populations shows substantial overlap, and (ii) it is variable across the participating databases.

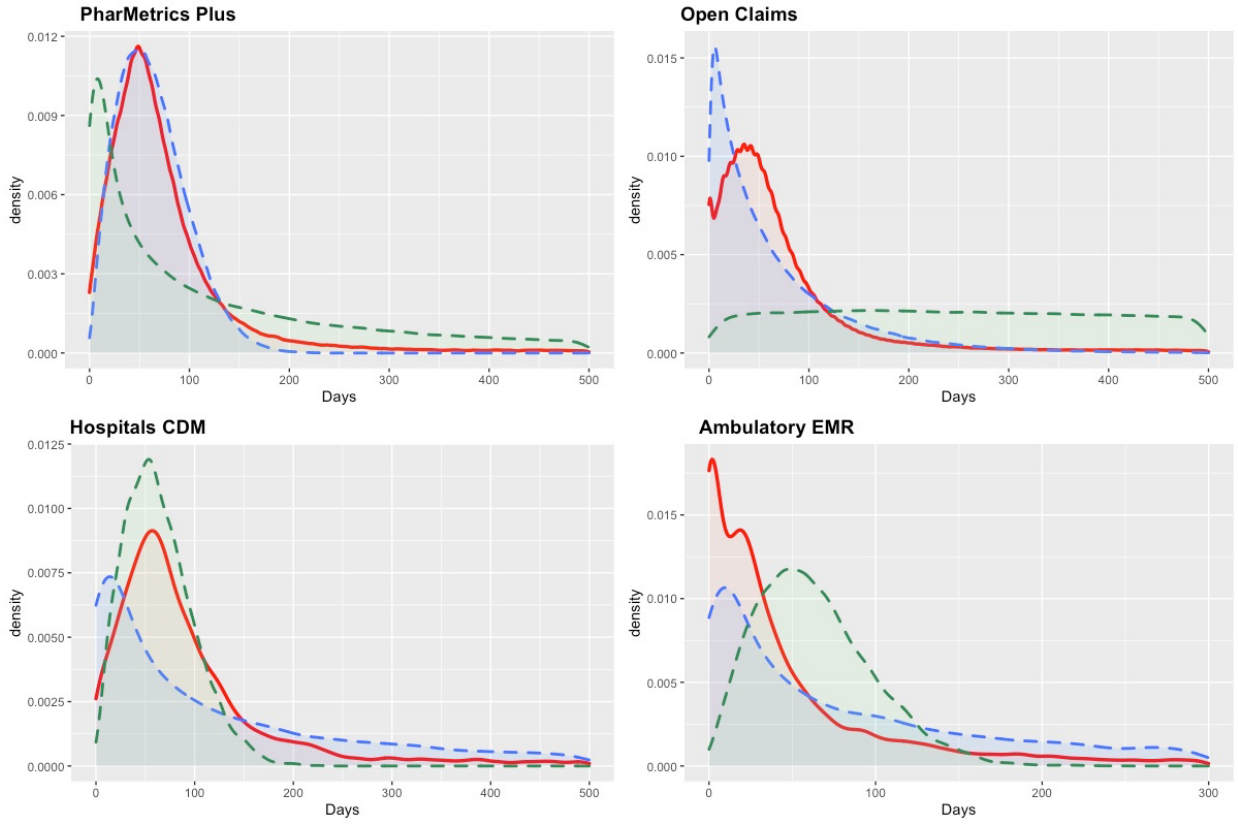


Fig 3. Empirical depiction of the two patient populations. Green and blue density plots represent the two distinct putative patient populations.

Conclusion

Curve fitting to finite mixture models via EM indicates that prostate cancer patients seem to be composed of two populations with different time to treatment characteristics, as expected from the treatment guidelines. The parameters of the optimal fitted models are in line with expectations: a mean between 49 and 76 days for the putative "immediate" group and 295 and 1067 days for the "deferred" group, a shape around +1 indicating a constant rate of TTT in both cohorts and a proportion with a dominant "immediate" group in the claims versus a 50/50 distribution in the ambulatory setting. The latter is consistent with the fact that most immediate treatments are initiated in the first or subsequent hospitalization, with no EHR records in an outpatient visit in between.

Further work is needed to develop a heuristic for an optimal TTT cut-off. However, even an optimal cut-off may not be practical for studying treatment intent in observational data given

these results: the variability between data and the apparent dependence on the context makes drawing generalizable conclusions a challenging task. In addition, the large overlap in the TTT distribution between putative "immediate" and "deferred" populations might cause a substantial degree of patient misclassification.