Assessing the Completeness of Oncology Treatment Data from Administrative Claims: A Benchmarking Study Against Abstracted EHRs Using Patient-Level Linkages

Joshuah Kapilivsky[1], Emma Roth[1], Zachary Rivers[1], Adam J. Hockenberry[1], Sandeep Jain[2], Jeremy L. Warner[2], Emilie Scherrer[1], Chithra Sangli[1]
[1]Tempus AI, Inc., Chicago, IL, USA, [2]Brown University, Providence, RI, USA

TEMPUS

LEGORRETA CANCER CENTER BROWN UNIVERSITY | Center for Clinical Cancer Informatics and Data Science

Poster Code: SA56

INTRODUCTION

Integrating real-world data (RWD) sources can enable increasingly granular precision oncology studies. Comparisons between data sources typically report aggregate statistics from unlinked datasets, but this approach precludes analysis of patient-specific date agreements. In this study, we leverage deterministic patient linkages to benchmark claims oncology treatment data against abstracted electronic health records (EHR) in a time-aware manner.

METHODS

We extracted abstracted EHRs from the Tempus multimodal database for 6487 stage 4 lung adenocarcinoma patients diagnosed between 2020 and 2023. Claims data (open and closed; medical and pharmacy) were linked using de-identified patient tokens. Claims between patients' first and last abstracted treatment dates were selected. Abstracted data were considered ground truth: claims for the same medication between abstracted start and end dates were true positives, unmatched claims false positives, and unmatched abstracted treatments false negatives. Abstracted treatment events were compared against closed claims if they took place during closed claims enrollment and compared against open claims otherwise.

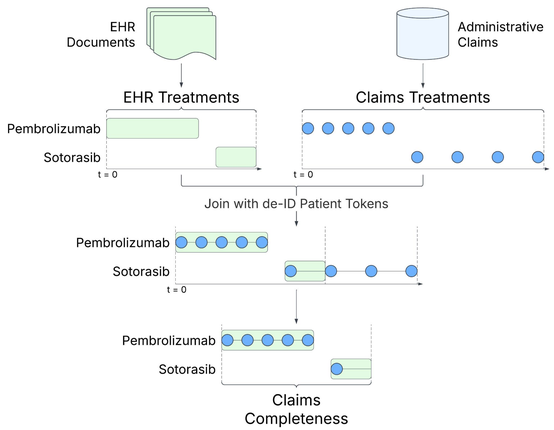

Figure 1. Treatment data from abstracted EHRs and administrative claims were combined at the patient level using deterministic patient linkages. Completeness metrics were calculated using the claims between patients' first and last abstracted treatment dates. Note that a single abstracted treatment period may correspond to one or many claims treatment events.

CONCLUSIONS

- PPVs for open and closed claims relative to EHRs indicate that individual claims may be sufficient to identify patient eligibility based on oncology treatment history for an RWD study.
- The sensitivities and start date match rates suggest closed claims may be suitable to extend comprehensive cancer treatment journeys beyond what is available from abstracted EHRs.

RESULTS

Schematics for calculating sensitivity and PPV of oncology treatment data from closed and open claims relative to abstracted EHRs

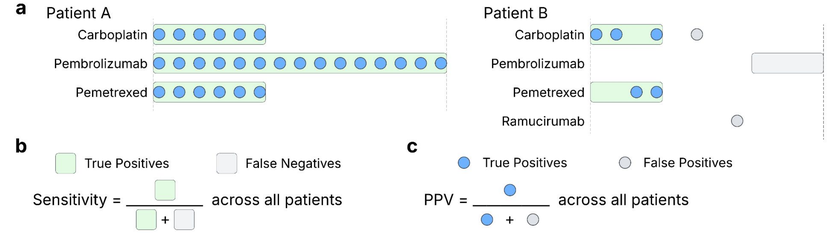

Figure 2. a) Each patient's treatment events from claims were compared against their treatment history as abstracted from EHRs. Only abstracted treatments with a known start and end date were included. Disagreements between sources are shown in gray. b,c) After classifying each event, sensitivity (b) and positive predictive values (PPVs, c) were calculated by aggregating across all events for all patients.

Sensitivities and PPVs for oncology treatment data from closed and open claims

		Closed Claims							Open Claims						
		Sensitivity				PPV			Sensitivity				PPV		
Treatment		True Positives	Total	%	TP | FN	True Positives	Total	%	TP | FP	True Positives	Total	%	TP | FN	True Positives	Total	%	TP | FP
All Treatments		***1235***	***1449***	***85.2***		***7384***	***8687***	***85.0***		***1928***	***4529***	***42.6***		***9540***	***11057***	***86.3***	
All Infusions		***1162***	***1339***	***86.8***		***6522***	***7716***	***84.5***		***1855***	***4289***	***43.3***		***8991***	***10485***	***85.8***	
Bevacizumab	Biologic	17	22	77.3		87	103	84.5		23	51	45.1		131	213	61.5	
Ramucirumab	Biologic	20	22	90.9		65	72	90.3		23	71	32.4		127	145	87.6	
Carboplatin	Chemotherapy	393	431	91.2		1495	1653	90.4		606	1400	43.3		2132	2315	92.1	
Cisplatin	Chemotherapy	40	48	83.3		128	135	94.8		66	140	47.1		193	197	98.0	
Docetaxel	Chemotherapy	24	26	92.3		134	154	87.0		38	111	34.2		191	227	84.1	
Gemcitabine	Chemotherapy	9	10	90.0		53	67	79.1		23	42	54.8		187	196	95.4	
Paclitaxel	Chemotherapy	76	87	87.4		362	457	79.2		114	276	41.3		600	711	84.4	
Pemetrexed	Chemotherapy	293	335	87.5		1662	2088	79.6		492	1061	46.4		2368	2858	82.9	
Atezolizumab	Immunotherapy	13	17	76.5		95	101	94.1		12	26	46.2		111	145	76.6	
Durvalumab	Immunotherapy	41	43	95.3		652	663	98.3		77	143	53.8		763	770	99.1	
Ipilimumab	Immunotherapy	17	21	81.0		33	39	84.6		21	56	37.5		66	86	76.7	
Nivolumab	Immunotherapy	22	28	78.6		112	121	92.6		33	79	41.8		222	323	68.7	
Pembrolizumab	Immunotherapy	197	249	79.1		1644	2063	79.7		327	833	39.3		1900	2299	82.6	
All Orals		***73***	***110***	***66.4***		***862***	***971***	***88.8***		***73***	***240***	***30.4***		***549***	***572***	***96.0***	
Alectinib	ALK Inhibitor	16	21	76.2		196	208	94.2		9	36	25.0		67	73	91.8	
Osimertinib	EGFR Inhibitor	49	73	67.1		606	692	87.6		60	176	34.1		457	472	96.8	
Sotorasib	KRAS Inhibitor	8	16	50.0		60	71	84.5		4	28	14.3		25	27	92.6	

Table 1. Closed claims enrollment periods (left) showed greater sensitivities than open claims (right). Sensitivities also differed by route of administration, with infusions higher than orals. Regardless of claim type, PPVs were high. Denominators reflect differences in treatment incidence, temporal alignment with closed claims enrollment periods, and the one-to-many relationship between abstracted treatments and individual claims events.

Schematics for calculating the availability of abstracted oncology treatment start and end dates in closed claims

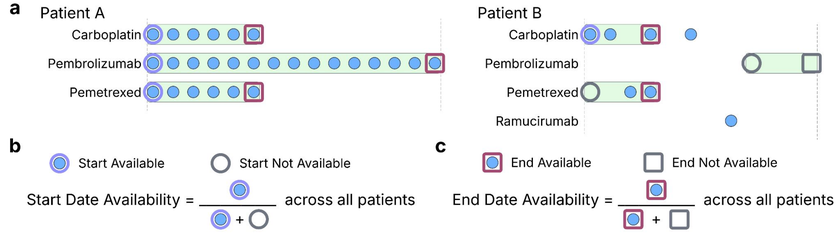

Figure 3. a) Claims data were searched for an event matching the abstracted start/end date per treatment. An exact date match was required for infusions, while a ±7 day tolerance was applied for orals to account for prescription fill delays. b,c) Availabilities were calculated by aggregating across all start/end dates for all patients. Treatments abstracted as ongoing were excluded from end date availability calculations.

Availabilities of abstracted oncology treatment start and end dates in closed claims

		Closed Claims							
		Start Date Availabilty				End Date Availability			
Treatment		Start Available	Total	%	Yes | No	End Available	Total	%	Yes | No
All Treatments		***1505***	***2168***	***69.4***		***663***	***1039***	***63.8***	
All Infusions		***1367***	***1905***	***71.8***		***654***	***997***	***65.6***	
Bevacizumab	Biologic	15	27	55.6		6	13	46.2	
Ramucirumab	Biologic	22	35	62.9		7	13	53.8	
Carboplatin	Chemotherapy	411	550	74.7		229	337	68.0	
Cisplatin	Chemotherapy	30	46	65.2		24	41	58.5	
Docetaxel	Chemotherapy	34	47	72.3		13	17	76.5	
Gemcitabine	Chemotherapy	16	22	72.7		6	7	85.7	
Paclitaxel	Chemotherapy	71	93	76.3		41	71	57.7	
Pemetrexed	Chemotherapy	361	487	74.1		175	249	70.3	
Atezolizumab	Immunotherapy	10	22	45.5		8	11	72.7	
Durvalumab	Immunotherapy	47	57	82.5		24	31	77.4	
Ipilimumab	Immunotherapy	25	32	78.1		5	12	41.7	
Nivolumab	Immunotherapy	31	42	73.8		11	21	52.4	
Pembrolizumab	Immunotherapy	294	445	66.1		105	174	60.3	
All Orals		***138***	***263***	***52.5***		***9***	***42***	***21.4***	
Alectinib	ALK Inhibitor	29	44	65.9		1	8	12.5	
Osimertinib	EGFR Inhibitor	101	190	53.2		7	28	25.0	
Sotorasib	KRAS Inhibitor	8	29	27.6		1	6	16.7	

Table 2. Abstracted start dates tended to have a matching claim at higher rates than abstracted end dates. The difference was more pronounced in orals, which generally had lower match rates than infusions.

Looking for more?
See our companion poster SA71 (Fri. 9 - 11:30 AM): "Integrating Next Generation Sequencing, EHR, and Claims Data to Extend Follow-Up in a Real-World Advanced Lung Adenocarcinoma Biomarker-Treatment Landscape" (Abstract number: 4214)

Correspondence: joshuah.kapilivsky@tempus.com