Integrating data from disparate sources to create a comprehensive patient journey: a case study in prostate cancer

Pooja Chopra[1], Wei Zhang[1], YiYuan Zhai[1], Sudha Sundarrajan[1], David Kang[1], Rebecca Garnett[2], Vasu Chandrasekaran[1], Nicholas J. Robert[1]
[1]Ontada, Irving, TX, USA [2]RTI International, Raleigh, NC, USA

Background

- Appropriate linkage of patient data from multiple sources can help uncover deeper insights into the patient journey to maximize potential of novel therapies or services[1, 2]
- The purpose of this study was to highlight the challenges of reliably and compliantly integrating different data sources as well as best practices. The study integrated electronic health record (EHR) data with open medical and pharmacy claims data, as well as social determinants of health (SDOH) data among patients with prostate cancer to obtain a more complete view of the disease burden and treatment journey.

Methods

Data Source

- iKnowMed™ (iKM) is an integrated web-based database of the oncology-specific EHR system maintained by McKesson Life Sciences. iKM captures outpatient practice encounter histories as well as demographics and clinical information.
- The third-party open claims data is a HIPAA-compliant, de-identified patient/ customer-level longitudinal database consisting of Medical and Pharmacy claims.
- Third-party social determinants of health (SDOH) data contain demographic and consumer information for a subset of patients

Study Population (see Figure 1)

Statistical Analyses

- Patient characteristics (Cohorts 1 and 2); social determinants (Cohort 3)
- Specialty visits in the pre-oncology period as well as post-oncology period: patients were evaluated by whether additional specialty visits could be captured from 3rd-party claims to fill the gap of iKM dataset (Cohort 2).
- Patients were also evaluated by whether treatments outside oncology clinics could be captured from 3rd-party claims (Cohort 2).

Results

- Comorbidity data were similar across data sources.
- Information including medical resource utilization, urology visit, and surgery was captured from the open claim data sources (**Table 1**).
- Third-party data showed that patients visit different types of specialties outside of oncology clinics **(Figure 2)**.
- Among Cohort 2, besides urology specialists, over half the patients had seen internal medicine, family medicine, and diagnostic radiology prior to oncology clinics.
- In the post-oncology period, the top 3 visited specialties were internal medicine, family medicine, and hematology/oncology.
- Additional treatment information outside of the oncology clinic setting was captured from the third-party open claim dataset (**Figure 3).**
- iKM overlapped with SDOH data (cohort 3) yields ~95-100% of the data for marital status, household income range, household size, language preference, and economic stability, and over 80% for education.

Figure 1. Study Population Selection Flowchart

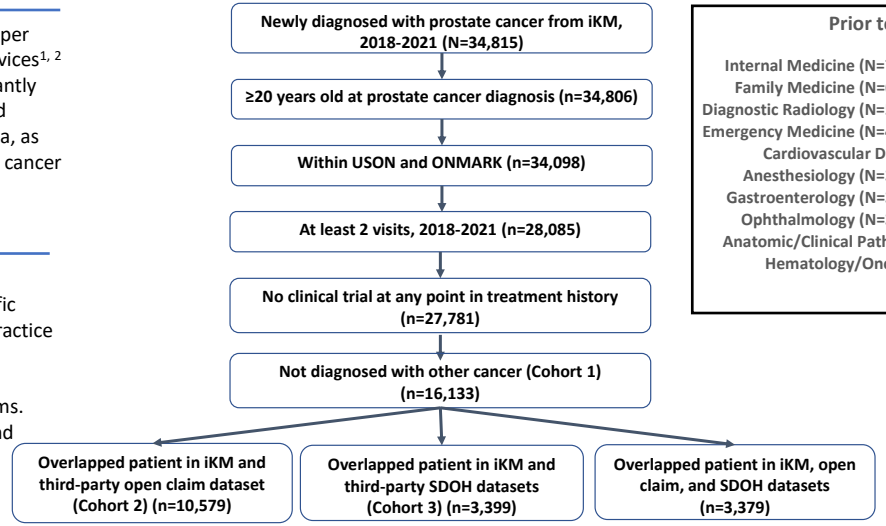

Figure 2. Evaluation of Top Ten Specialty Visits (Cohort 2; n=10,579)

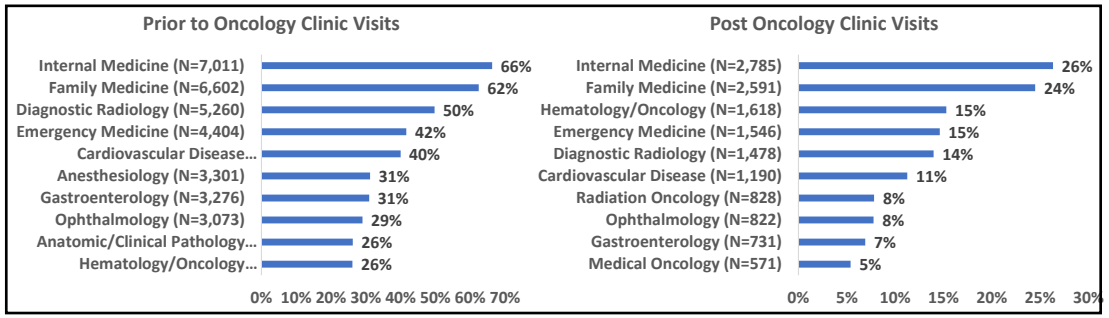

Figure 3. Evaluation of Treatments (Cohort 2; n=10,579)

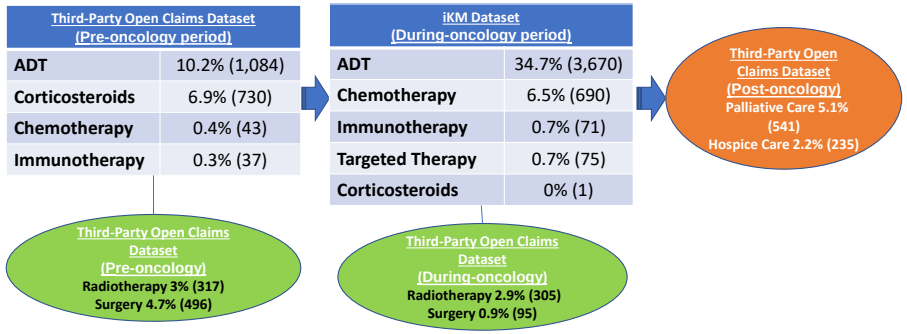

Table 1. Patient Characteristics across Data Sources

Patient Characteristics	iKM Data Only (N=16,133)	iKM Plus Open Claims (N=16,133)	iKM Overlapped Open Claims (N=10,579)
Age at diagnosis – mean (SD), years	71.6 (9.4)	71.6 (9.4)	71.9 (9.4)
PET/CT Scan - no. (%)	840 (5.2)	1,399 (8.7)	1,132 (10.7)
Bone Scan - no. (%)	709 (4.4)	3,777(23.4)	3,609(34.1)
MRI - no. (%)	1,058 (6.6)	1,202(7.5)	904 (8.5)
PSA Testing- no. (%)	13,062 (81.0)	13,739 (85.2)	9,701 (91.7)
Testosterone Level- no. (%)	4,624 (28.7)	5,380 (33.3)	1,316 (39.7)
Radiotherapy- no. (%)	14 (0.1)	2,054 (12.7)	2,052 (19.4)
Surgery - no. (%)			
Biopsy		30 (0.2)	30 (0.2)
Surgical Castration		4 (0.0)	4 (0.0)
Laparoscopic Prostatectomy		886 (5.5)	886 (5.5)
Open Radical Prostatectomy		35 (0.2)	35 (0.2)
Medical resource utilization - no. (%)			
Inpatient		3,008 (18.6)	3,008 (28.4)
Outpatient		7,426 (46.0)	7,419 (70.1)
Emergency department		801 (5.0)	800 (7.6)
Urology visit- no. (%)		6308 (39.1)	6,305 (59.6)
Laboratory		5,678 (35.2)	5,675 (53.6)
Hospice		5 (0)	5 (0)
Other		23 (0.1)	23 (0.2)

Discussion

- In this study, lack of information regarding marital status, household income range, highest education, occupation, medical resource utilization, urology visit, and surgery could be filled by the third-party datasets as a supplemental resource.
- Also, addition of the third-party datasets can provide a broader view of patients' journey for specialty visits, imaging, and treatments received inside and outside of oncology clinics.
- This study confirmed that the richness of patients' information can be improved by linking the three different types of electronic healthcare data sources: iKM data provides more granular detail of patients' clinical characteristics, the third-party open claim data has a broad view of patients' activities, and the third-party SDOH data provides patients' socioeconomic status.

References

1.Thompson CA, et al. eGEMs 2015;3(1):1127
2. West SL et al. Health Informatics J. 2014;20(1):22-34. doi: 10.1177/1460458213476506.