

Completeness of Source-level Data Items Based on Type of Source

AR Kahn , MJ Schymura , TM Juster

Study Objectives

- To identify the percent unknown values for key data items.
- To determine the level of missing information based on report source.
- To assess the percent unknown values by type of tumor.
- To analyze the level of concordance of source reports to consolidated information.

The extent of unknowns based on type of reporting source is important information as central registries become increasingly dependent on receiving reports from non-traditional sources, in order to maintain complete case ascertainment.

Methods

- Identify all sources records (n=646,384) that contribute to all reportable tumors (n=413,966) in the New York State Cancer Registry (NYSCR) for diagnosis years 2012 – 2014, as of May, 2016.
- Exclude non-NY residents, Death Certificate Only, Autopsy-Only, Out-of-state, VA hospital, and pediatric Early Case Capture records and tumors identified based on those sources. (resulting n=545,416 sources and n=353,969 tumors)
- Link each source record to the corresponding consolidated tumor data.
- Analyze the distribution of sources and consolidated tumors by type, specifically: Hospital inpatient or overall Managed Care Organization as Inpatient; Radiation-Oncology Center, Ambulatory Surgery Center, other Outpatient facility as Outpatient; Private physician office; or Laboratory (using the prescribed hierarchy as per NAACCR item #500 to arrive at the type for consolidated tumors). We combined the sources from Ambulatory Surgery Centers, Radiation Oncology Centers and Outpatient facilities because we found that there was a lot of overlap in these categories, especially in this current fluid healthcare environment.
- Compare the percent unknown of various data items among the raw sources; and among the consolidated tumors; and among both raw sources and consolidated values by type of tumor. Analysis of stage, subsite, and nodes was limited to types of tumors for which those items are meaningful.
- Use Excel for 'quick and dirty' linear regressions plotting the percent unknown race and node data against percent best source from inpatient and MD records, with one data point for each of 23 tumor types.
- Identify concordance between source values and consolidated values by type of source.

Overall Distribution by Reporting Source Type

Fig 1A

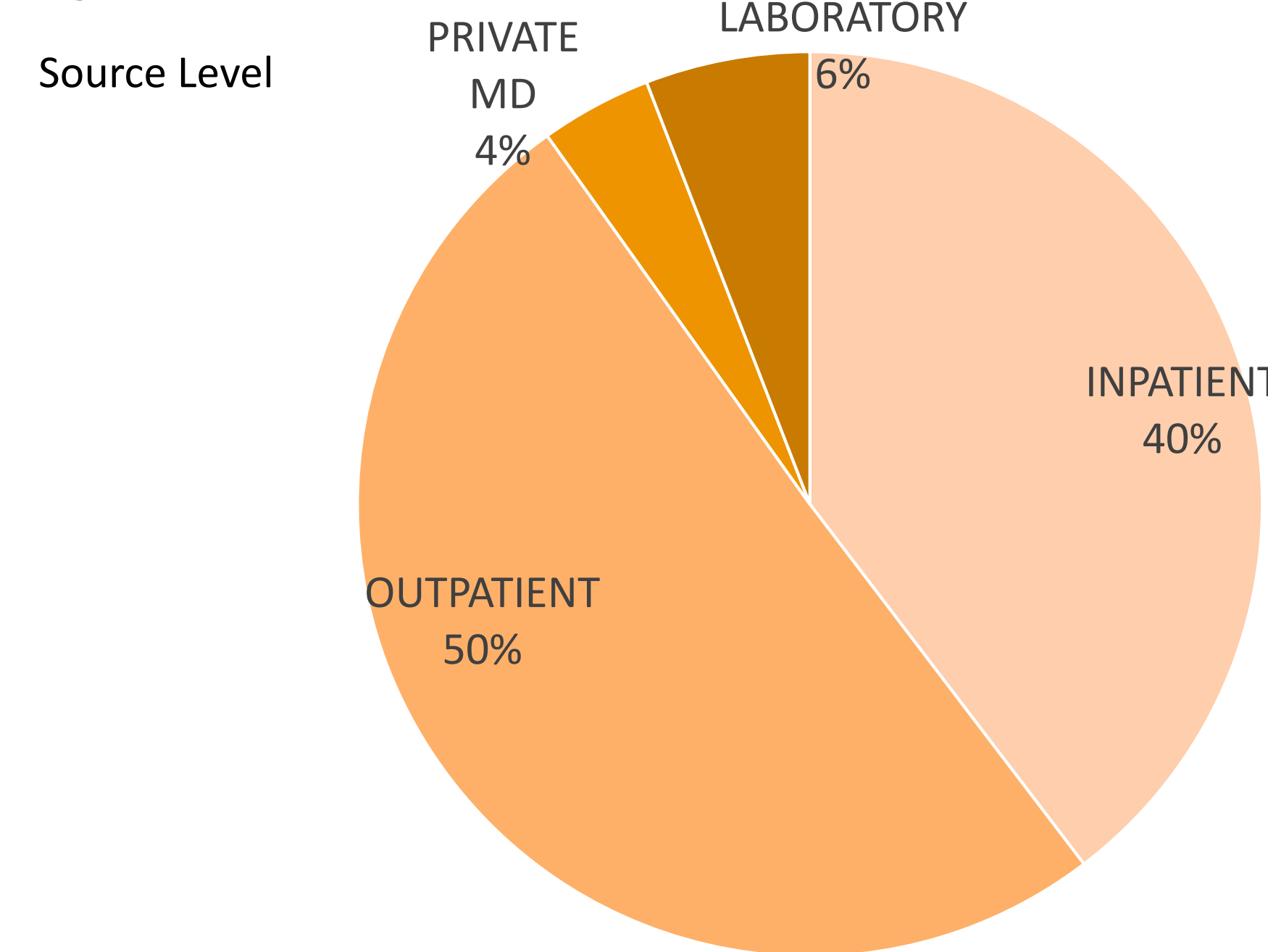
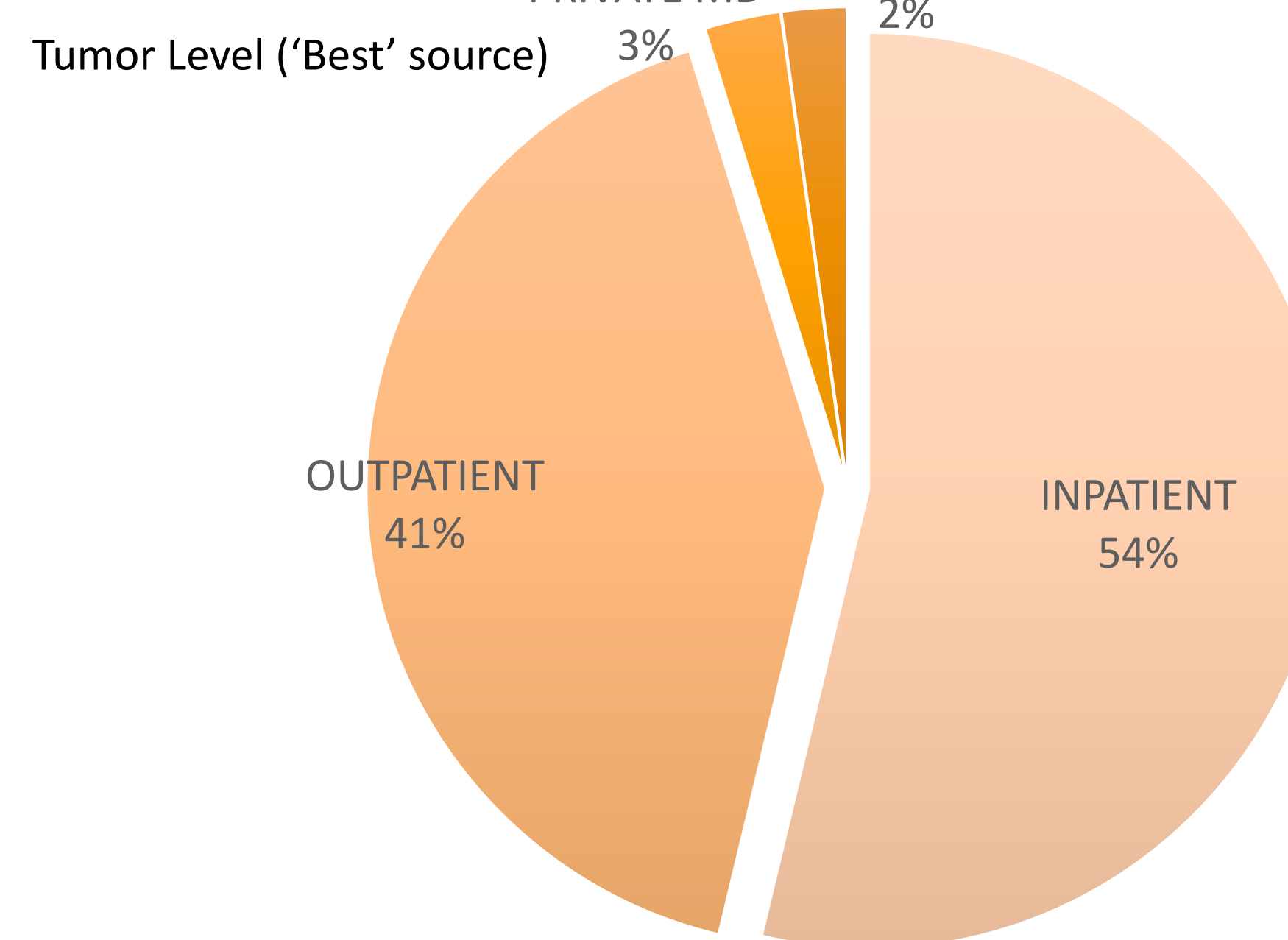


Fig 1B



% Unknown Values for Specific Items by Source Type

Table 1A Source Level					Table 1B Tumor Level ('Best' source)				
Item	Inpatient	Outpatient	Private MD	Lab	Item	InPatient	Outpatient	Private MD	Lab
Laterality	6.2	2.4	4.9	7.0	Laterality	2.8	1.2	4.3	8.8
Confirm.	1.7	0.7	0.4	0.0	Confirm	0.6	0.3	0.2	0.0
Grade	42.7	34.8	50.2	49.9	Grade	36.9	33.6	71.1	66.8
Race	2.6	4.5	19.7	58.6	Race	0.5	0.9	4.3	19.2
Hisp.Eth	4.2	8.9	32.0	60.5	Hisp.Eth	1.6	4.3	24.4	55.8
SS2000	11.7	19.6	46.6	51.3	SS2000	6.1	6.8	22.1	45.6
Subsite	25.7	21.9	28.6	34.0	Subsite	19.2	17.1	11.0	28.3
NodesPos	14.9	20.3	48.0	38.8	NodesPos	10.6	9.3	40.6	21.9
NodesEx	14.8	20.2	47.9	38.7	NodesEx	10.6	9.2	40.5	21.8

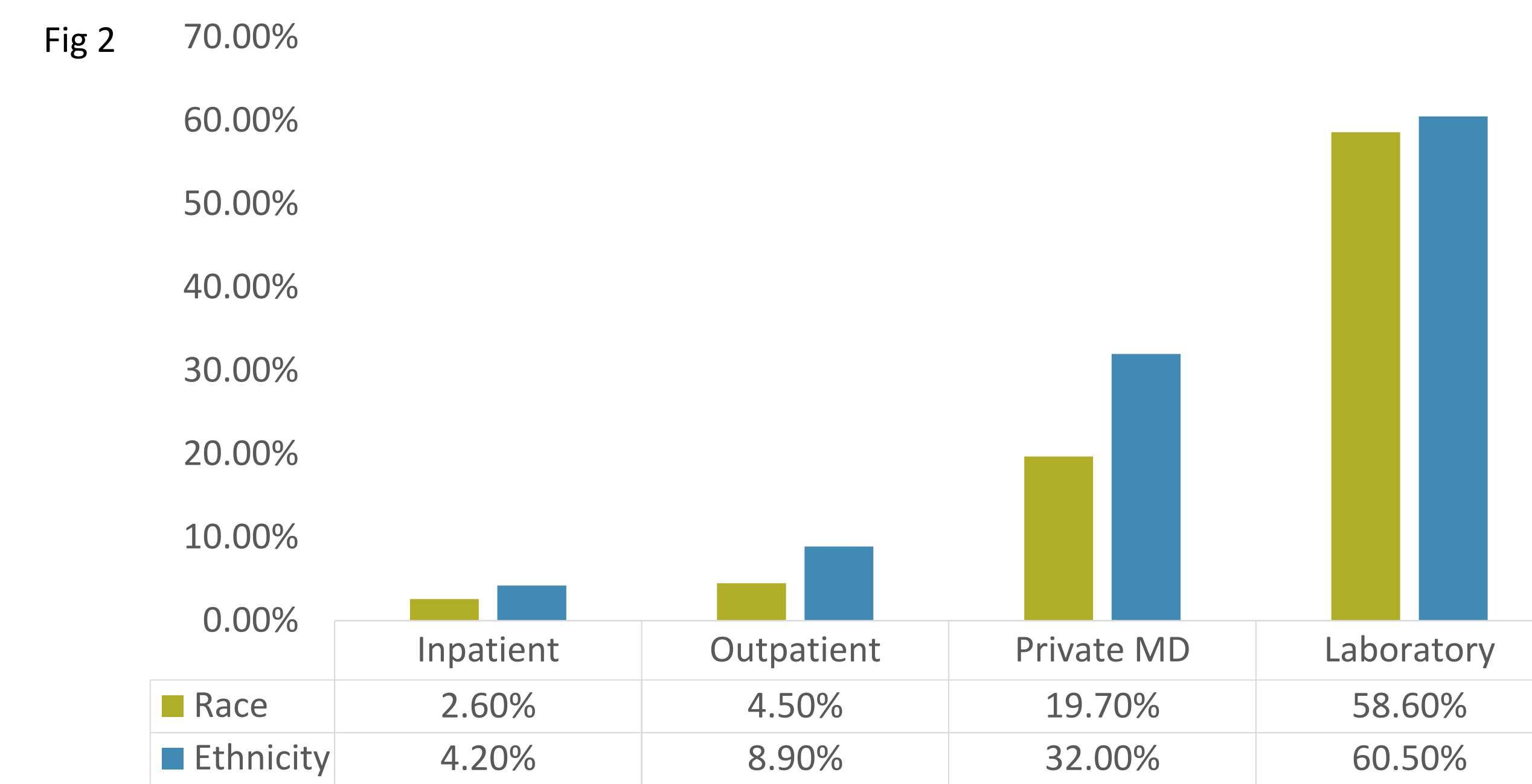
Type of Source for Different Tumors

Table 2A Source Level					Table 2B Tumor Level ('Best' source)				
Cancer	InP*	Rad_Onc/Amb_Surg	Private MD	Lab	Cancer	InP*	Rad_Onc/Amb_Surg	Private MD	Lab
Kidney	77.0	21.2	0.5	1.4	Kidney	87.8	12.0	0.1	0.1
CR	56.6	38.0	1.8	3.7	CR	76.9	21.4	0.7	1.1
Lung	54.0	43.2	0.6	2.2	Lung	74.4	25.2	0.1	0.3
Leukemia	53.0	32.5	4.5	10.0	Leukemia	59.4	28.5	5.1	7.1
Bladder	30.6	61.5	2.9	5.1	Bladder	36.5	60.2	1.7	1.6
Prostate	24.2	60.2	9.3	6.4	Prostate	34.1	60.0	3.3	2.6
Breast	19.4	71.7	1.8	7.1	Breast	32.1	66.4	0.5	1.0
Melanoma	7.4	43.8	27.6	21.2	Melanoma	9.4	48.9	28.2	13.5

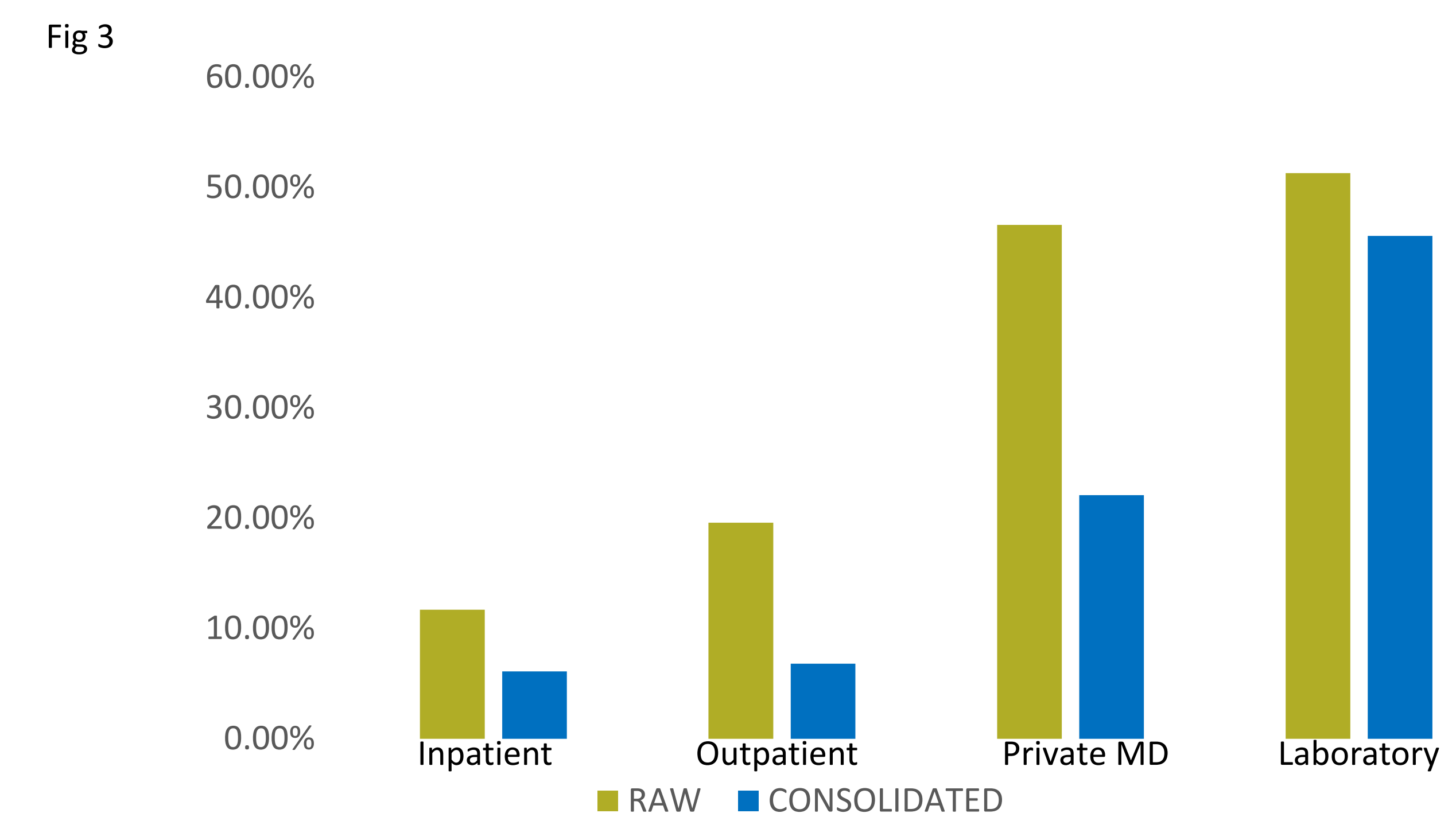
% Unknown Values for Specific Items by Diagnosis

Table 3A Source Level					Table 3B Tumor Level ('Best' source)				
Cancer	Race	Hisp.Eth	SS2000	Nodes+	Cancer	Race	Hisp.Eth	SS2000	Nodes+
Kidney	3.7	5.4	9.0	6.9	Kidney	0.8	2.2	3.2	1.5
CR	5.9	9.0	20.7	15.5	CR	0.9	3.0	6.3	3.6
Lung	3.6	7.3	18.2	16.9	Lung	0.3	1.5	4.7	3.4
Leukemia	9.7	12.2	NA	NA	Leukemia	2.2	7.3	NA	NA
Bladder	5.6	8.4	10.1	7.6	Bladder	0.5	3.1	3.8	2.2
Prostate	9.2	17.9	27.1	26.6	Prostate	1.9	10.3	14.8	15.6
Breast	8.2	11.8	16.0	14.0	Breast	0.9	2.6	2.2	1.6
Melanoma	19.9	26.4	15.4	14.6	Melanoma	4.0	17.7	9.2	8.5

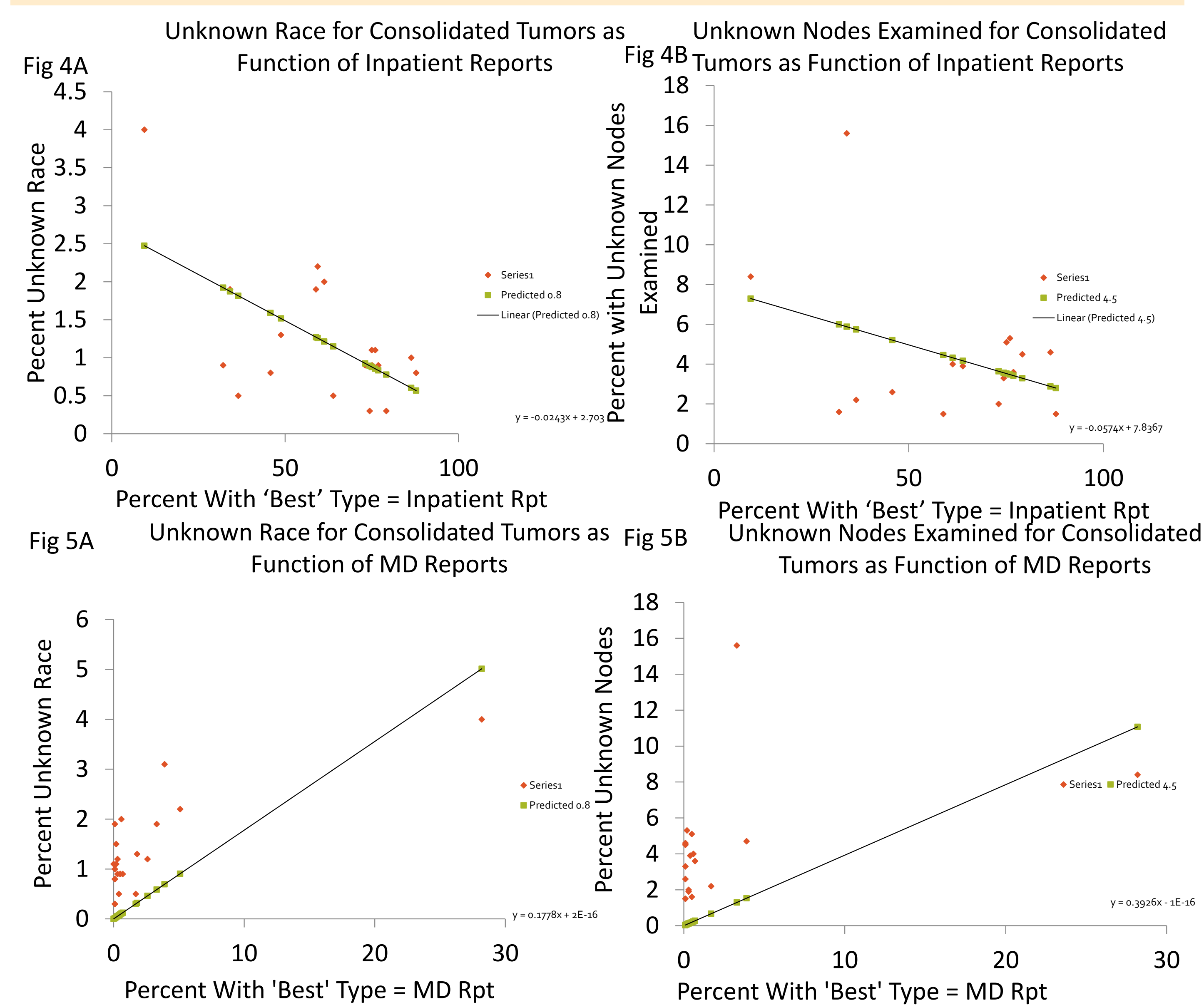
% Unknown Race and Ethnicity by Source Type



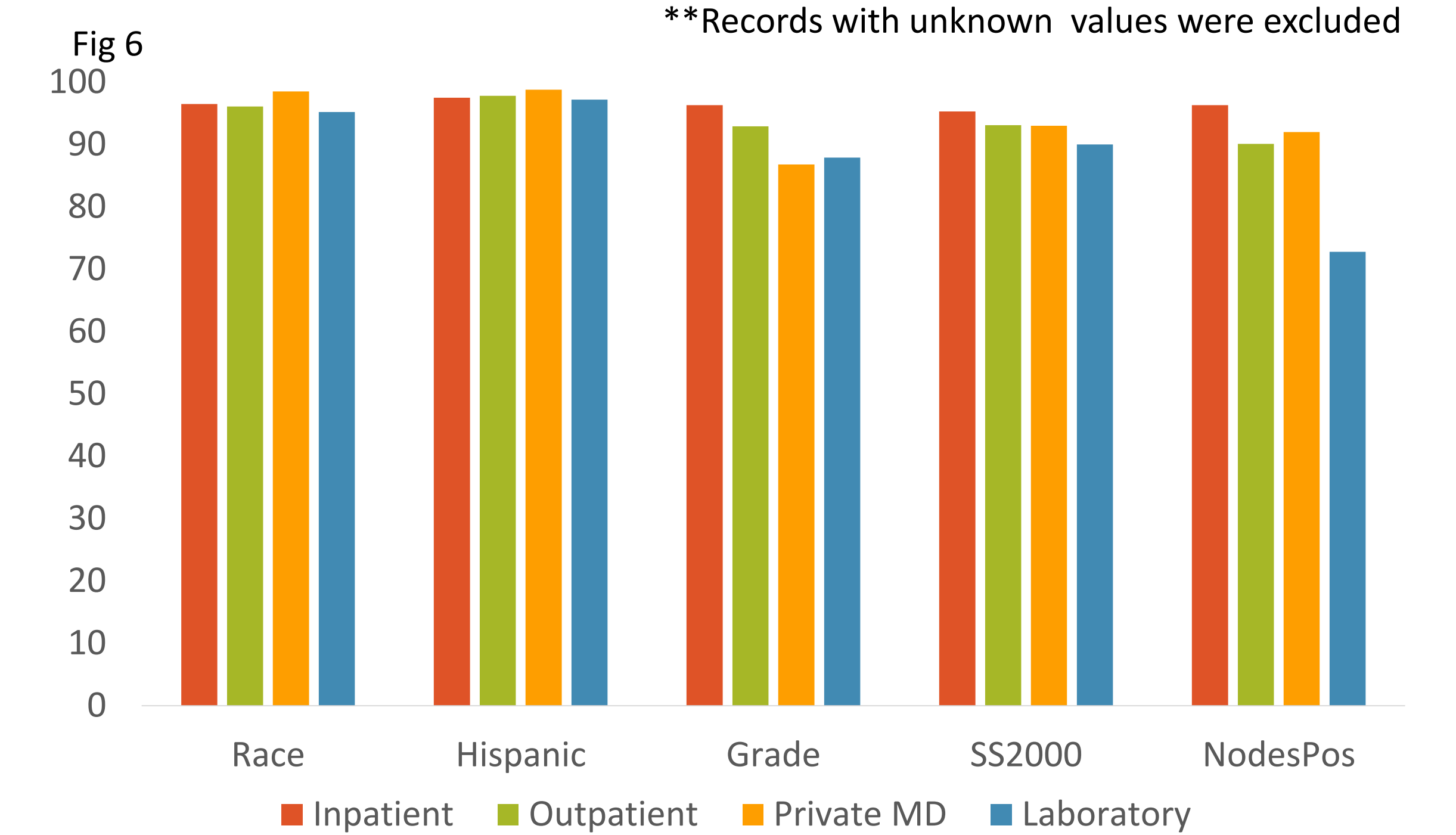
% Unknown Summary Stage Raw Source Type vs 'Best' Source Type



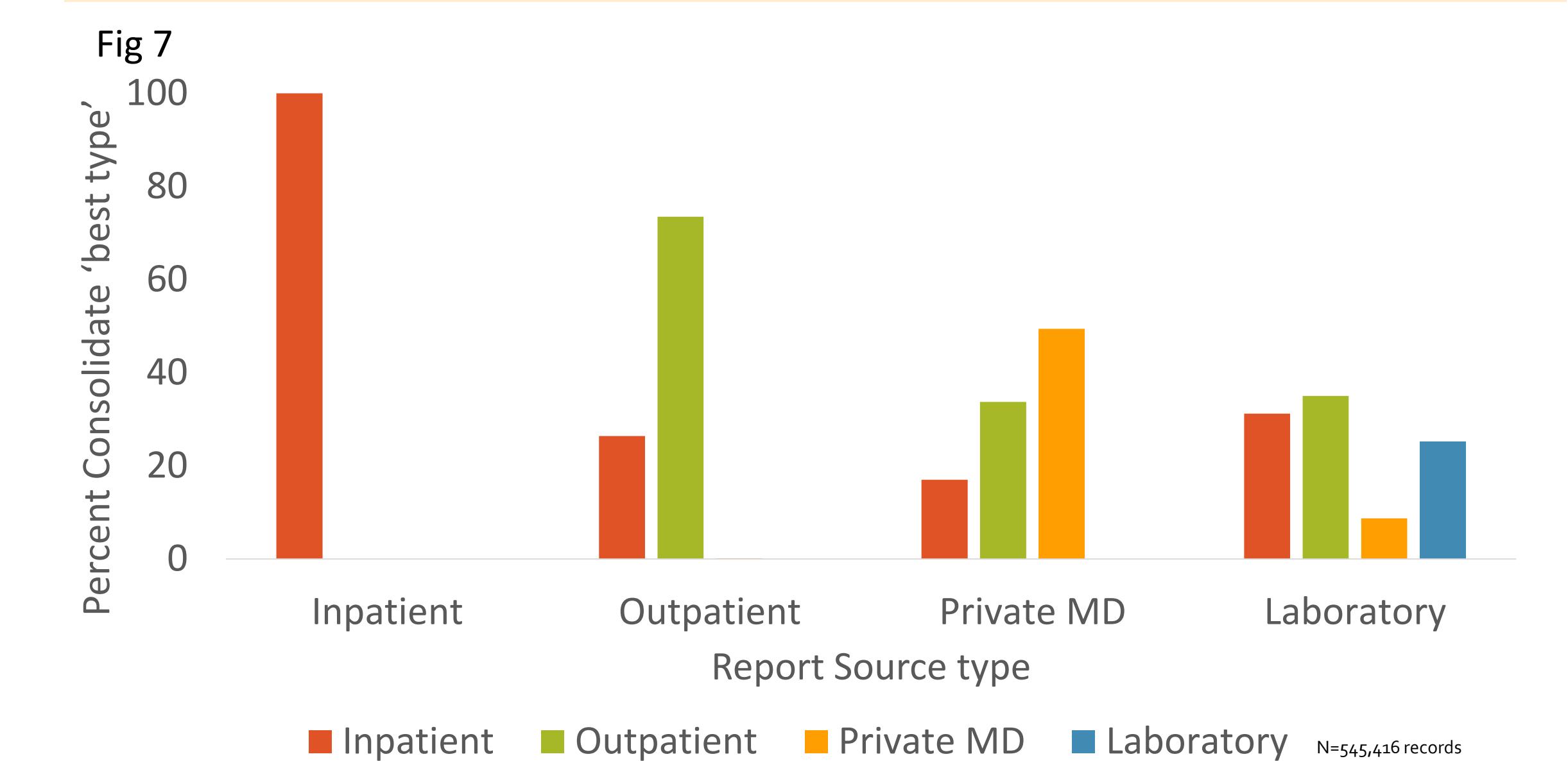
Correlation of 'Best' Source With Unknown Values, including all Diagnoses



% Concordance of Source Values With Consolidated Values**



Best Source Type by Original Source



Findings

- The New York State Cancer Registry (NYSCR) receives more reports from outpatient sources (Radiation Oncology Centers, Ambulatory Surgery Centers, or Outpatient Services of Hospitals) than from any other type of facility.
- However, the NYSCR receives an inpatient report for a majority of tumors (best source for 54% of the tumors was an Inpatient Hospital report). Reports from physicians' offices and laboratories accounted for 3.6% and 6% of the source records but only 2.6% and 2.2%, respectively, of the best sources for consolidated tumors.
- Records from laboratories and physicians' offices had the highest percent of unknown values for race and Hispanic ethnicity. These sources were also missing information on regional lymph node status and summary stage.
- Records from outpatient facilities were similar in percent unknown to inpatient records. Outpatient records had more complete information than inpatient records for laterality, diagnostic confirmation, and grade, probably associated with breast cases.
- The source of reports varies greatly by site. Whereas 77% of kidney cancer reports come from Inpatient records, only 7.4% of melanoma reports are from hospital inpatient records.
- Simple linear regressions plotting the percent unknown race and unknown lymph nodes show inverse relationships with percent inpatient' best source' type and direct relationships with percent MD 'best source' type.
- Concordance of known source values with their corresponding consolidated tumor values, by source type, for race, Hispanic ethnicity, grade, summary stage, and number of positive lymph nodes show high agreement of sources with consolidated values for inpatient, outpatient, and private physician's records, for known values.
- For 100% of tumors with an inpatient report, the 'best type' is inpatient. Almost 30% of outpatient reports also had an inpatient report, but for almost 50% of MD reports and 25% of lab reports, there were no corresponding Inpatient or Outpatient reports.

Conclusions

More than half of the tumors reported to the NYSCR have at least one Inpatient report. However, the type of report source varies based on diagnosis, with almost ninety percent of kidney cancers but only nine percent of melanomas having an Inpatient report. Records from Laboratories and from Private MDs have more unknown values than other records. It is reassuring that records from different sources with known values generally agree with the consolidated tumor values. A challenge facing central registries is to maintain a low percent of unknown values, particularly for race and ethnicity, for tumors that are only reported by Laboratories or Private MDs.