Cancer Care Ontario
Action Cancer Ontario

Quality Control of Alternate Data Sources in the Ontario Cancer Registry

Mary Jane King
Cancer Care Ontario
NAACCR 2015, Charlotte, North Carolina

Ontario
Cancer Care Ontario
Action Cancer Ontario

Background

- Cancer Care Ontario (CCO) decommissioned the Ontario Cancer Registry Information System (OCRIS, cases 1964-2009) in 2014
- It was replaced with the new Ontario Cancer Registry (OCR, cases 2010 onward)
- This is the first significant revision of Ontario's cancer registry rules and technology since the 1980s
- OCRIS case counts were based on an Ontario-specific variant of IARC multiple primary rules
- OCR now conforms to the SEER MPH rules for counting multiple primaries

Impetus for Review & Correction I

- Inherent need for review to account for the differences in case counting rules between OCRIS and the new OCR
  - Develop techniques to mitigate change when doing trend analysis both OCRIS and OCR data
  - Best practices for presenting annual incidence rates, survival and prevalence*

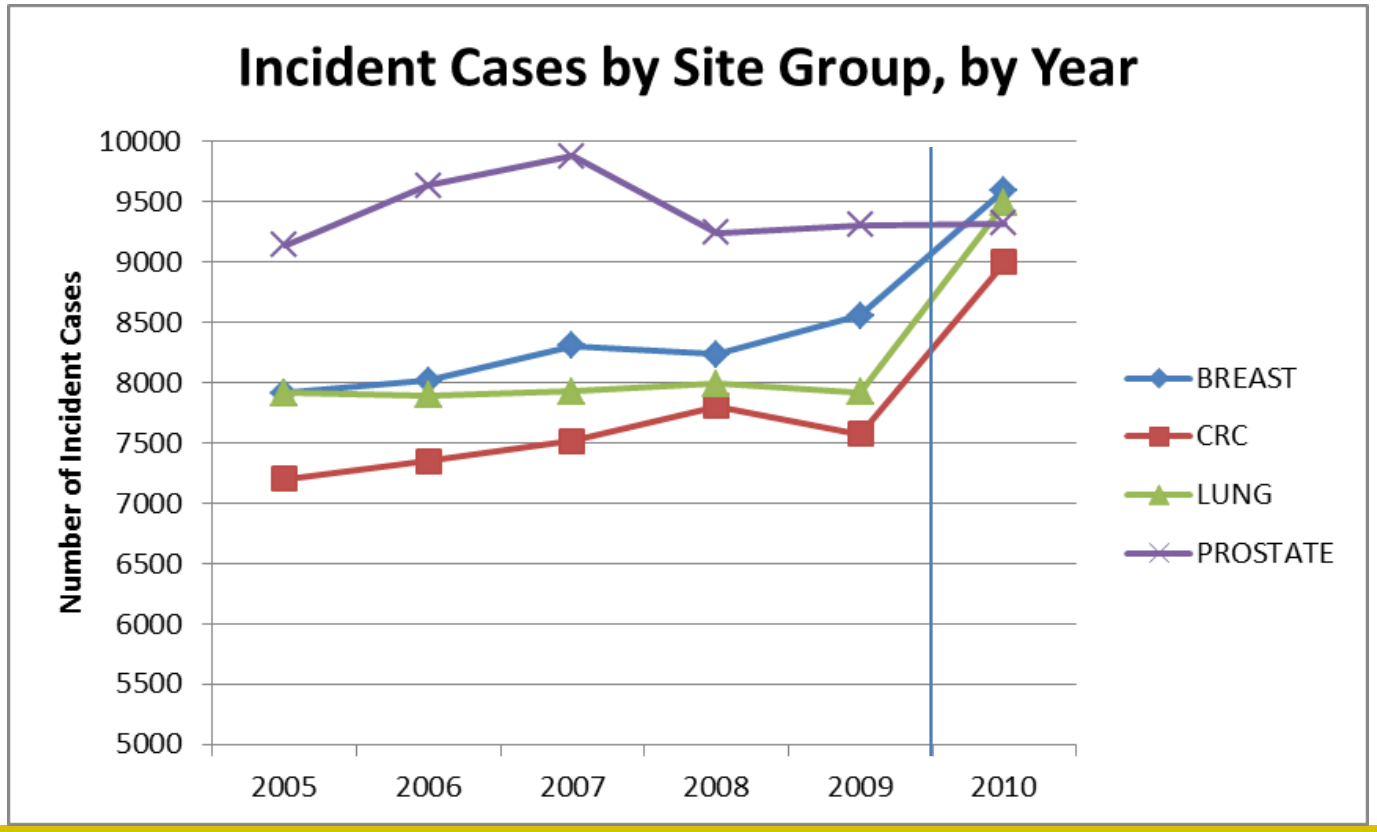
- Provide a Business Impact Assessment (BIA) and FAQs for data analysts and consumers of OCR reports and statistics
  - CCO Clinical & Regional Programs
  - Cancer System Quality Index for Cancer Quality Council of Ontario
  - Surveillance, Research, Calls for Data

* See Appendix

OCRIS to OCR

Why has reported number of cases increased?

- More liberal rules for multiple primary cancer sites
- Additional source records
- Records that were ignored before are now being included



Impetus for Review & Correction II

- OCR cases are still created from re-purposed administrative data sources
    - Different types of source records, different coding standards, different standards of “truth”
- Conservative OCRIS “undercounting rules” masked some of these effects
- OCRIS standards for the “credibility” of source records was modified for the new OCR, but not enough
    - OCRIS over counted cases – gave incident status to cases created solely from inpatient discharge data
- OCR was built for intervention
    - OCR can be manually corrected at the case level; OCRIS could not

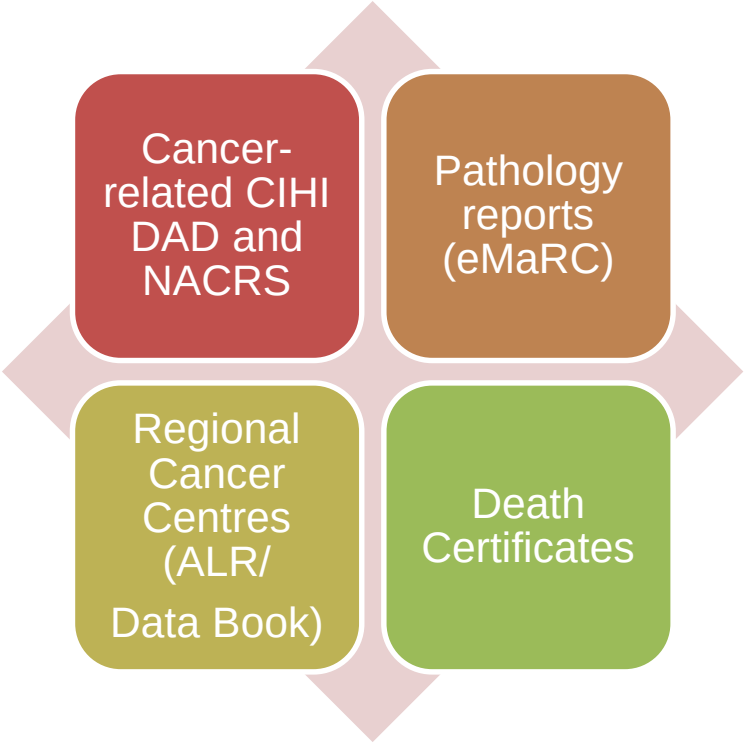
OCR Description & Sources

A computerized database of information about all Ontario residents who have been diagnosed with cancer ("Incidence") or who have died of cancer ("Mortality")

Must support collection of cancer incidence without the support of a hospital cancer registry system

A surveillance registry embedded in a healthcare funding agency

4 Major Data Sources



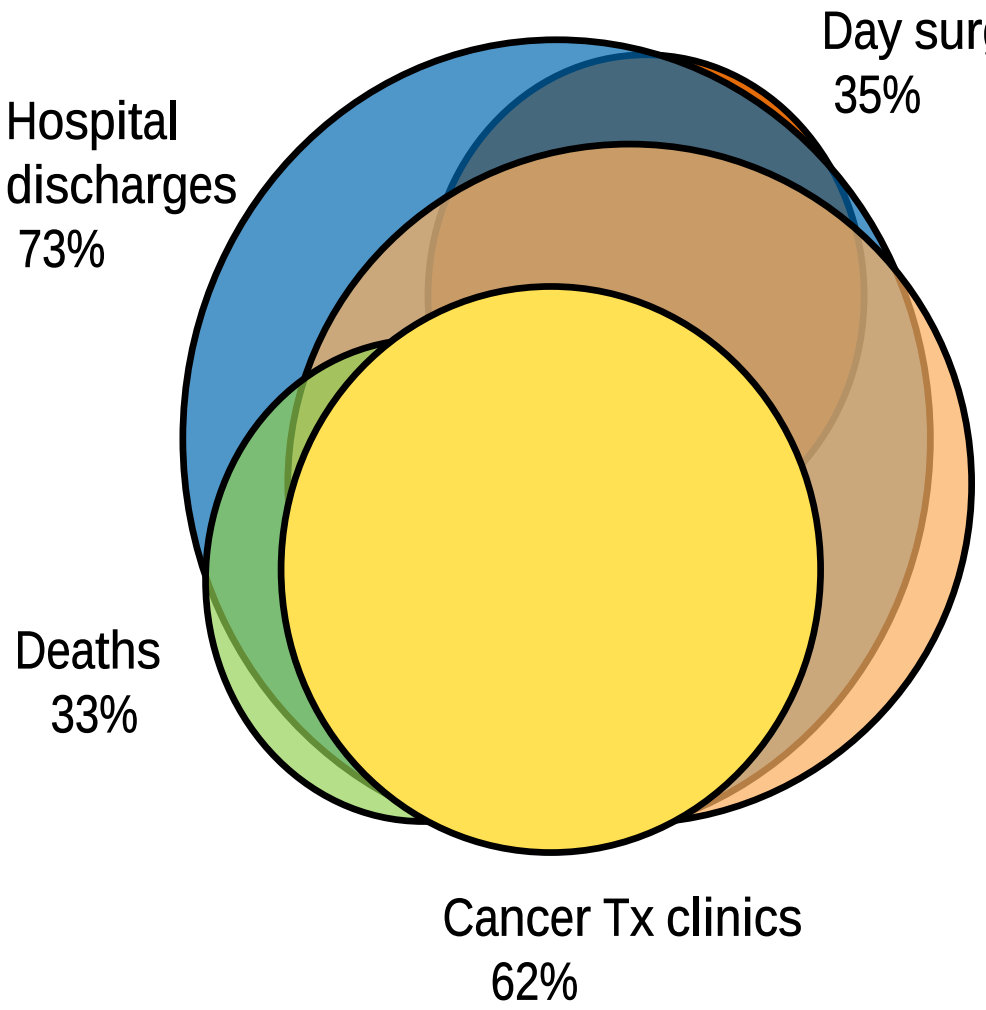
OCR Key Automated Processes

- **Patient Linkage**
  - Combination of deterministic and probabilistic linkage routines to aggregate a person's source records (what was submitted to CCO) into a "best of linked person record"
  - Generates a single composite/representative record representing that individual

- **Case Resolution**
  - Associates a Person's records from multiple data sources into one or more primary cases of cancer
  - Generates a single "resolved" record representing each primary case

Source Accrual Patterns & Adjustment

Reporting sources for 2007 incident cases in OCRIS, Oct. 2013



% Single Source	
Hosp only	4%
Path only	5%
DC only	2%
Clinics only	3%
(NACRS day surg	2%)

OCR Adjustments

~~Hosp only~~

~~Laterality singletons~~

~~ALR singletons~~

Removing Hospital Only Cases

- OCRIS inpatient discharge only cases (DAD) – incident
- OCRIS outpatient discharge only cases (NACRS) – no case

- OCR DAD only cases – Problematic (non-incident, not credible)
- OCR NACRS only cases – also Problematic

- Plan - that the more lethal cancers would be exempted from the OCR DAD=Problematic. Subsequently, most were made Problematic
- This was proven to be a mistake for Lung due to serendipity
    - Half of 2012 CS lung cases came from OCRIS, half from OCR
    - A significant number of OCRIS DAD only cases could be staged!

Source accrual patterns for Lung versus Breast

Source accrual – Percentage of incident **lung** cases with one or more of each source

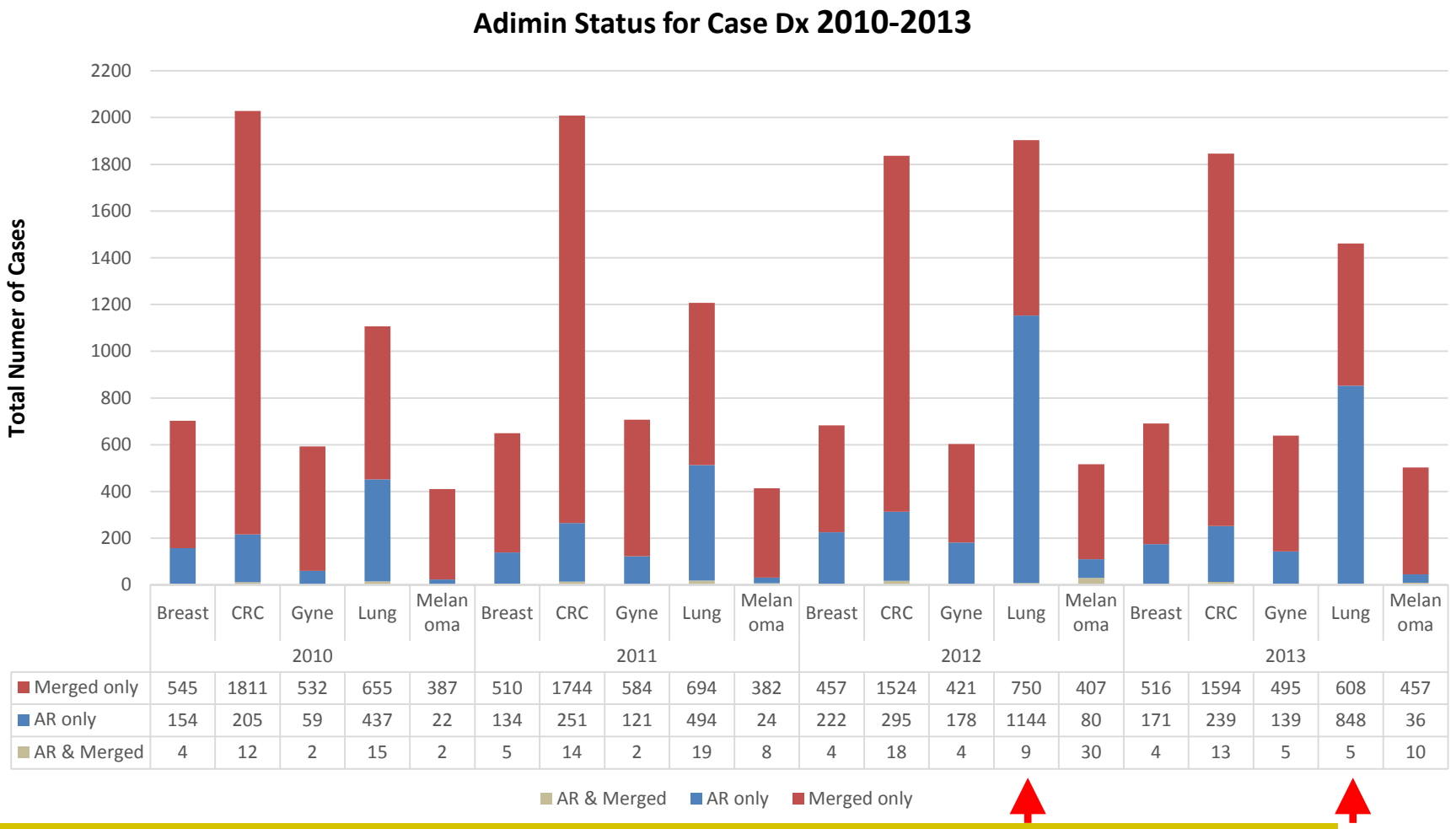
Source	2010	2011	2012
NACRS	68.4	68.5	67.9
DAD	80.6	79.6	79.3
Path	74.9	76.9	77.1
ALR	64.5	65.3	65.3
Death	60.5	52.9	33.8

Source accrual – Percentage of incident **breast** cases with one or more of each source

Source	2010	2011	2012
NACRS	81.5	81.3	82.0
DAD	45.4	41.9	40.0
Path	95.5	96.5	96.6
ALR	80.5	80.9	83.6
Death	7.8	5.0	2.8

Manual Review of OCR

Administrative Changes¥ & Merging False Positives



	2010					2011					2012					2013				
	Breast	CRC	Gyne	Lung	Melanoma	Breast	CRC	Gyne	Lung	Melanoma	Breast	CRC	Gyne	Lung	Melanoma	Breast	CRC	Gyne	Lung	Melanoma
Merged only	545	1811	532	655	387	510	1744	584	694	382	457	1524	421	750	407	516	1594	495	608	457
AR only	154	205	59	437	22	134	251	121	494	24	222	295	178	1144	80	171	239	139	848	36
AR & Merged	4	12	2	15	2	5	14	2	19	8	4	18	4	9	30	4	13	5	5	10

¥ Mainly changing Problematic to Incident

Some Exceptions to Discharge = Problematic

- Leukemia has always been the cancer with the highest % hospital only / lowest % microscopic confirmation because we don't get the peripheral blood smears / bone marrow reports.
    - Any rule about how to treat hospital onlies has the greatest effect of leukemia
    - Currently all hematopoietic cancers are exempt and MP are manually reviewed

- All CNS tumours are exempt and manually reviewed

Source Records with Generic/Variant Coding

- CIHI DAD and NACRS use ICD-10-CA
  - C509 converts to C509 8000/3
  - CIHI most relevant diagnosis field needs investigating

- Cancer Centre/ALR can use ICD-10 or ICD-O-3
  - ICD-O-3 often placeholder quality
  - **C509 8140/3 is perfectly acceptable for treatment funding purposes**

- ePath autocoding is dependent on words for narrative reports
  - **If final diagnosis “breast carcinoma” or “adenocarcinoma” = C509 8010/3 or 8140/3**

MPH Logic Changes

Example - Site pairs too restrictive, solid tumours

- Adenocarcinoma NOS should stay with other more specific adenocarcinomas
- Incorrectly resolved to two cases - resolve to one case of the further histology down in the histologic tree
    - For most buckets, the rule that says “Tumors with ICD-O-3 histology codes that are different at the first (Xxxx), second (xXxx) or third (xxXx) number are multiple primaries” needs to be modified. e.g. see Breast rule H12
    - Also, Rule M11 is too restrictive and should be modified to allow source records with certain histology groupings to associate (e.g. 8140/3 with more specific adenocarcinomas). Once these records are associated, H3 rule pointing to histology pairs tables will ensure that the correct (more specific) histology is selected

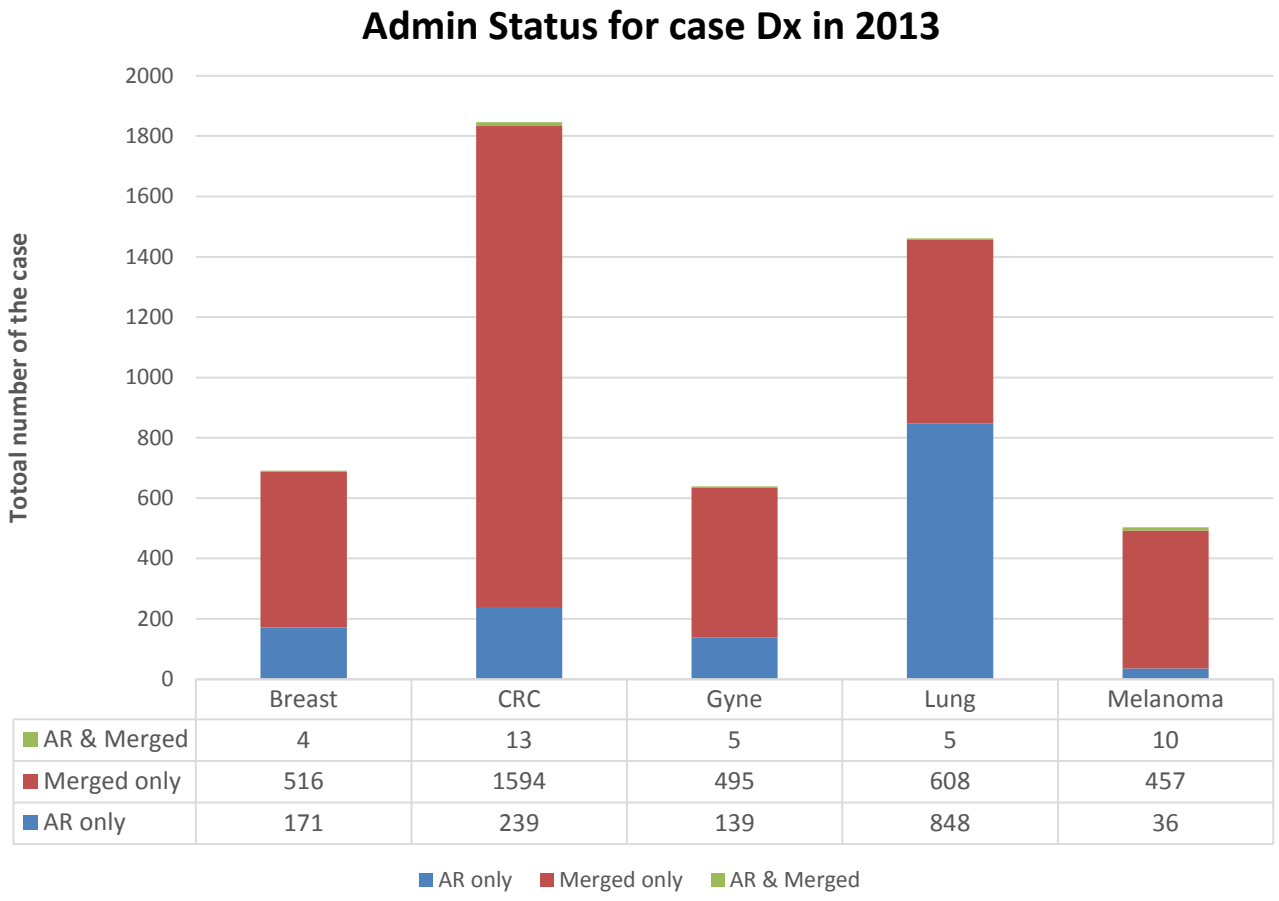
Breast Rules H12, M11, H3

H12	Code the most specific histologic term when the diagnosis is: • Carcinoma, NOS (8010) and a more specific carcinoma or • Adenocarcinoma, NOS (8140) and a more specific adenocarcinoma or • Duct carcinoma, NOS (8500) and a more specific duct carcinoma (8022, 8035, 8501-8508) or • Sarcoma, NOS (8800) and a more specific sarcoma Note: The specific histology may be identified as type, subtype, predominantly, with features of, major, with ___ differentiation. The terms architecture and pattern are subtypes only for in situ cancer.	IF One histology is 8010, 8140, 8500, or 8800 AND The other histology is in the table of histologies related to the first histology THEN Take the more specific histology ELSE continue processing with H26	This rule will be executed when both records have behavior /3 (invasive) and the hierarchy can't rank one above the other This rule uses the following tables: Carcinoma NOS and more specific carcinomas; Adenocarcinoma NOS and more specific adenocarcinomas; Duct carcinoma and more specific duct carcinomas; Sarcoma NOS and more specific sarcomas The tables of related histologies should include the NOS terms themselves (e.g. the 8010 table should contain 8010) so that e.g. two 8010 histologies would satisfy this rule.

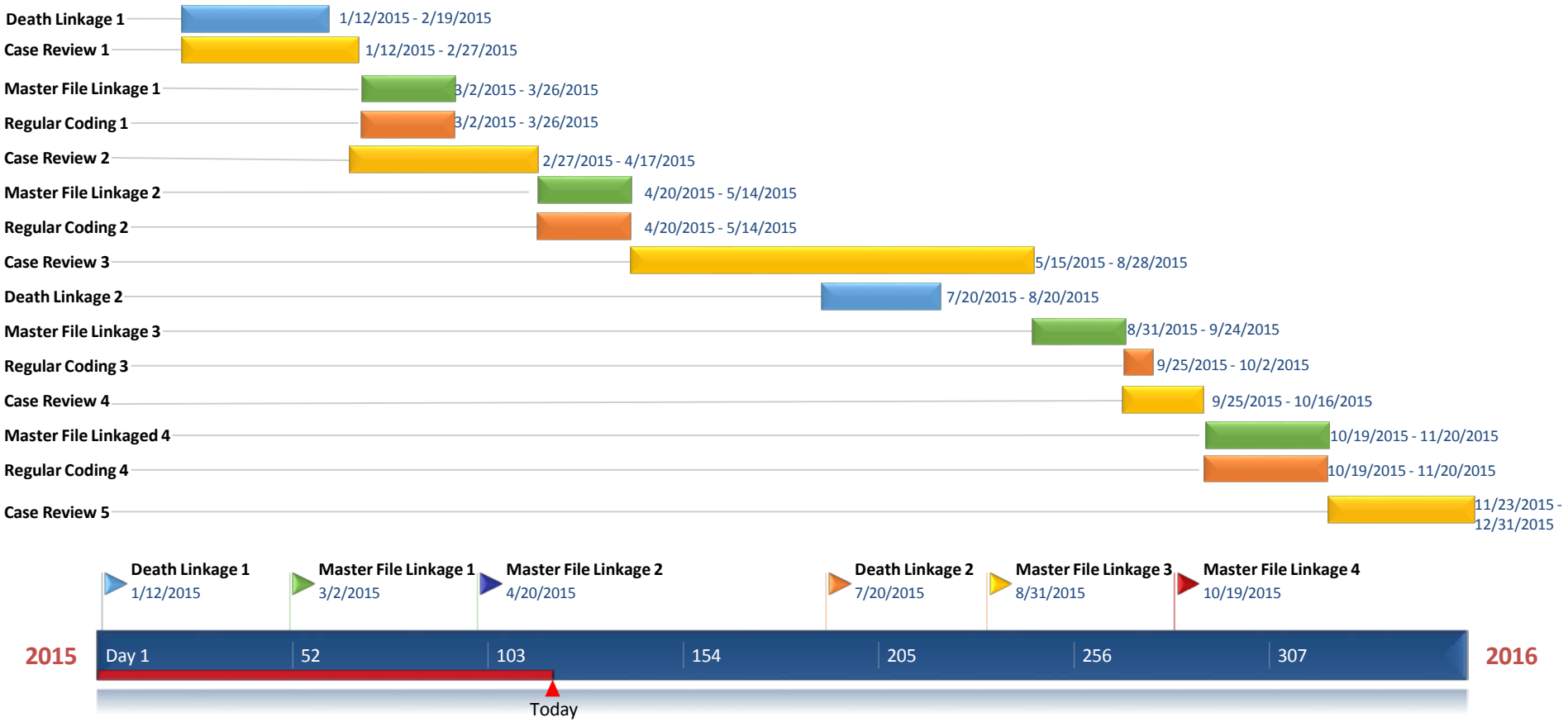
M11	Multiple intraductal and/or duct carcinomas are a single primary. Note: Use Table 1 and Table 2 to identify intraductal and duct carcinomas	IF The histology on each record in the table of duct carcinomas or the table of intraductal carcinomas THEN The records are ASSOCIATED ELSE Continue processing with next rule

H3	Code the more specific histologic term when the diagnosis is: • Carcinoma in situ, NOS (8010) and a specific carcinoma in situ or • Adenocarcinoma in situ, NOS (8140) and a specific adenocarcinoma in situ or • Intraductal carcinoma, NOS (8500) and a specific intraductal carcinoma (Table 1) Note: The specific histology may be identified as type, subtype, predominantly, with features of, major, with ___ differentiation, architecture or pattern. The terms architecture and pattern are subtypes only for in situ cancer.	IF One histology is 8010, 8140, or 8500 AND The other histology is in the table of histologies related to the first histology THEN Take the more specific histology ELSE continue processing with next rule	This rule will be executed when both records have behavior /2 (in situ) and the hierarchy can't rank one above the other This rule uses the following tables: Carcinoma NOS and more specific carcinomas; Adenocarcinoma NOS and more specific adenocarcinomas; Intraductal and specific intraductal carcinomas, which contains the values {8201, 8230, 8401, 8500, 8501, 8503, 8504, 8507} The tables of related histologies should include the NOS terms themselves (e.g. the 8010 table should contain 8010) so that e.g. two 8010 histologies would satisfy this rule

Site	AR only	Merged only	AR & Merged	Total
Breast	171	516	4	691
CRC	239	1594	13	1846
Gyne	139	495	5	639
Lung	848	608	5	1461
Melanoma	36	457	10	503
Grant Total	1433	3670	37	5140



OCR Linkage/Case Res Schedule alternating with Manual Pathology Coding

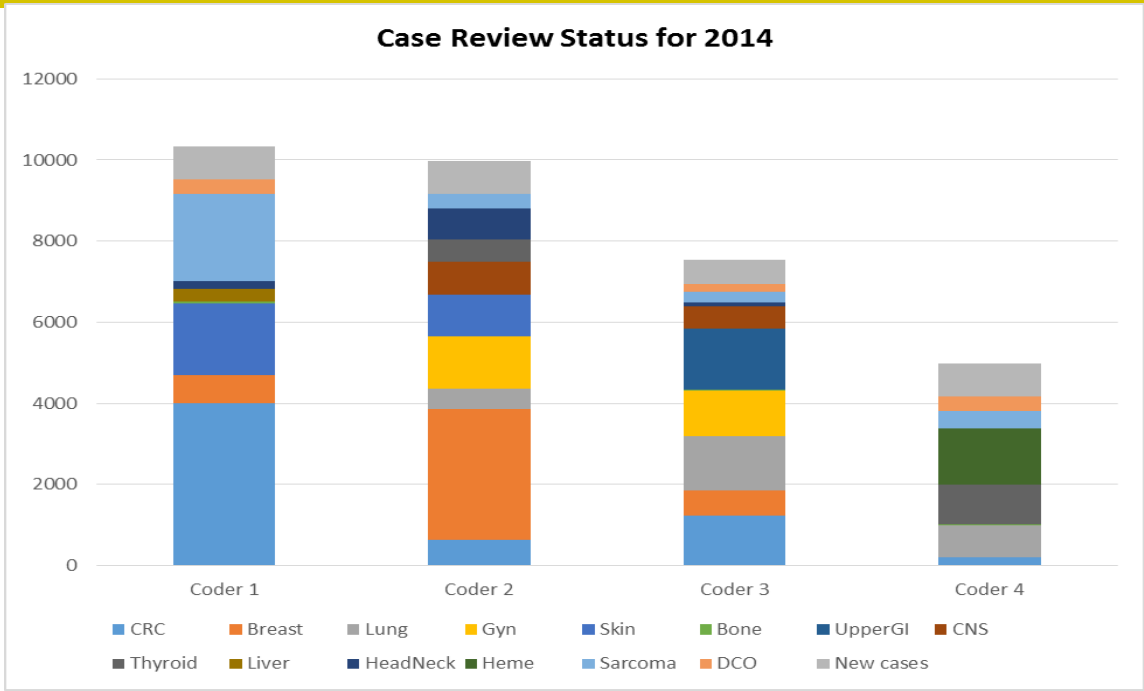


OCR remediation takes place in a busy environment throughout the year, including Master Linkage, Death Linkage

It competes with other processes for resource time, such as “Regular Coding” – manual coding of path reports

There is a race to provide CS staging lists of merged, correct cases once every quarter

Manual Review of Multiple Primaries and Hem Cases



Coder	Coder 1	Coder 2	Coder 3	Coder 4	Grant Total
CRC	4011	628	1238	202	6079
Breast	688	3225	626	0	4539
Lung	0	510	1327	795	2632
Gyn	0	1285	1126	0	2411
Skin	1768	1042	0	0	2810
Bone	35	0	19	17	71
UpperGI	0	0	1517	0	1517
CNS	0	793	528	0	1321
Thyroid	0	568	0	983	1551
Liver	318	0	0	0	318
HeadNeck	202	760	101	0	1063
Heme	0	0	0	1376	1376
Sarcoma	2145	355	265	449	3214
DCO	350	0	200	346	896
New cases	808	808	604	804	3024
Grant Total	10325	9974	7551	4972	32822

Acknowledgements & Contacts

Diane Nishri

Senior Scientist, Prevention & Cancer Control, Population Health

Karen Hofmann

Senior Analyst, OCR - Data Analytics and Person Linkage

Grace Liu

Team Lead OCR - Pathology coding, OCR manual review, Linkage

Mary Jane King

Manager, OCR

ocrquestions@cancercare.on.ca

Appendix: Acronyms

CCO – Cancer Care Ontario

OCRIS – Ontario Cancer Registry Information System

OCR – refers here to the new automated Ontario Cancer Registry system

CIHI – Canadian Institute for Health Information

DAD – Discharge Abstract Database (inpatient) - CIHI

NACRS – National Ambulatory Reporting System (outpatient) - CIHI

Appendix: *Best Analytic Practices

- When analyzing incidence trends that include both OCRIS and OCR data, it is recommended that only cases that meet the IACR Multiple Primary rules be included.
- For all other analyses, all appropriate primaries should be included
  - Annual incidence rates – include all primaries
  - Survival – analyze first primary per cancer
  - Prevalence – analyze either first primary per person, or first primary per cancer, or first primary per person per time period

Appendix: Example OCR Analysis & FAQ - Breast

-Diane Nishri

Microsoft Word Document