Non-Traditional Sources to Support Cancer Surveillance NAACCR 2020

Lynne Penberthy
NCI/ SEER

NIH NATIONAL CANCER INSTITUTE

Objectives

- Briefly discuss the use of non-traditional data sources to support cancer surveillance & why this is important
- Describe several examples of the use of these sources by SEER

Pace of Change in Cancer Care

- How cancer patients are diagnosed and managed is changing at a pace that is so rapid it is almost impossible even for clinicians to keep up to date
- There are no data outside clinical trials that provide information on
  - How these new diagnostic methods and treatment are used and
  - How they impact outcomes

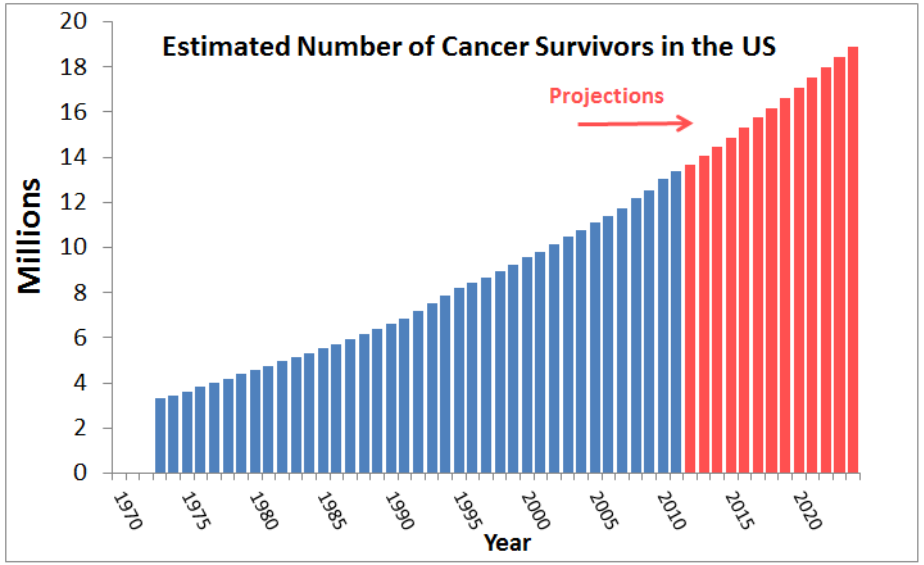
Why is this critical to the cancer surveillance mission?

- Guidelines for treatment are based on clinical trials but…
  - These capture <5% of the cancer population (often fewer than 20 patients)
  - Are largely biased- white and no comorbidities
  - We need population level data to understand the use and effect of these new treatments in patients not represented in clinical trials

Why is this critical to the cancer surveillance mission?

Example: Understanding recurrence risk

- With nearly 17 million cancer survivors in the US alone (nearly 5% of the population) it is essential that we provide these data to patients & clinicians
- Many clinical trials are now focused on recurrent disease and our most intransigent cancers with the highest mortality are likely to manifest with recurrent metastatic disease
  - Pancreas
  - Ovarian
  - Melanoma
  - GBM



[1] DeSantis C, Chunchieh L, Mariotto AB, et al. (2014). Cancer Treatment and Survivorship Statistics, 2014. CA: A Cancer Journal for Clinicians. In press.

How is SEER using non-traditional sources to address some of these gaps?

Non-traditional methods for surveillance being used by SEER

- Developing linkages with many external partners who hold key data for understanding cancer care
- Developing methods to automate data capture through deep learning and NLP
- Beginning to leverage new technologies for enhancing information
  - Digital pathology images
  - Radiology reports/images
  - Liquid biopsies

Linkages

Linkages for treatment

- Pharmacy data
  - Pharmacists are health care providers and thus covered under most state regulations
  - Thousands of pharmacies- therefore working central organizations for linkage
  - There are no population level data on the use and effect of oral anti-neoplastic agents in the US
  - SEER now linked with CVS/Walgreens/Riteaid across 15 registries (accounts for 50+ % of prescriptions in US)
    - “near real time reporting” to identify and track dissemination of new agents – and outcomes

Anti-neoplastic agent prescriptions from Walgreens and RiteAid - 2013-mid 2020

- >4 million prescriptions for anti-neoplastic drugs in SEER regions
- Includes agents & fills to monitor compliance
- Receiving monthly data submission
- Awaiting CVS full SEER submission
- Critical for more comprehensive capture of hormonal therapy

Data from Walgreens /RiteAid 2013-mid 2020	
	N
All	4,245,608
METHOTREXATE	1,601,097
ANASTROZOLE	777,312
TAMOXIFEN	548,756
HYDROXYUREA	183,641
MERCAPTOPURINE	163,712
PREDNISONE	139,768
EXEMESTANE	134,556
BICALUTAMIDE	122,224
LETROZOLE	106,693
LEUCOVORIN	82,279
ONDANSETRON	48,631
CAPECITABINE	31,452
LEUPROLIDE	19,727
METHYLPREDNISOLONE	18,782
FINASTERIDE	14,238
IBRANCE	13,864
GLEEVEC	13,776
CYCLOPHOSPHAMIDE	13,675
TEMOZOLOMIDE	13,551
XELODA	12,052
DEXAMETHASONE	10,405
SPRYCEL	9,191
ZYTIGA	9,016
LUPRON	8,883
LEUKERAN	8,816
ESTRADIOL	7,191
JAKAFI	7,154
TARCEVA	7,015
AFINITOR	6,989
POMALYST	6,384
TREXALL	6,328
ARIMIDEX	5,327
FEMARA	4,870
TASIGNA	4,634

Tyrosine Kinase Inhibitor Use among patients in SEER Reported through Walgreens/RiteAid (2013-mid 2020

Drug Name	Walgreens Number of Patients	Walgreens Number of Fills
ALECENSA	8	37
BOSULIF	83	625
CABOMETYX	113	618
Gilotrif	89	406
Gleevec	1196	13776
IMATINIB MESYLATE	2	2
INLYTA	206	1021
IRESSA	6	20
Iclusig	7	28
Lenvima	44	114
Nexavar	878	3585
RYDAPT	65	220
SPRYCEL	1016	9191
SUTENT	466	2215
Stivarga	464	1392
TAGRISSO	129	931
TARCEVA	934	7015
TYKERB	216	1000
Tasigna	444	4634
VOTRIENT	601	2754
Xalkori	96	737
ZYKADIA	19	94
Total	**7082**	**50415**

Frequency of TKI use from 2013-mid 2020. Provides an opportunity to:

- Monitor dissemination over time
- Evaluate outcomes among the general population rather than a limited CT population
- Understand compliance with prescribed treatment (using fills and intervals)

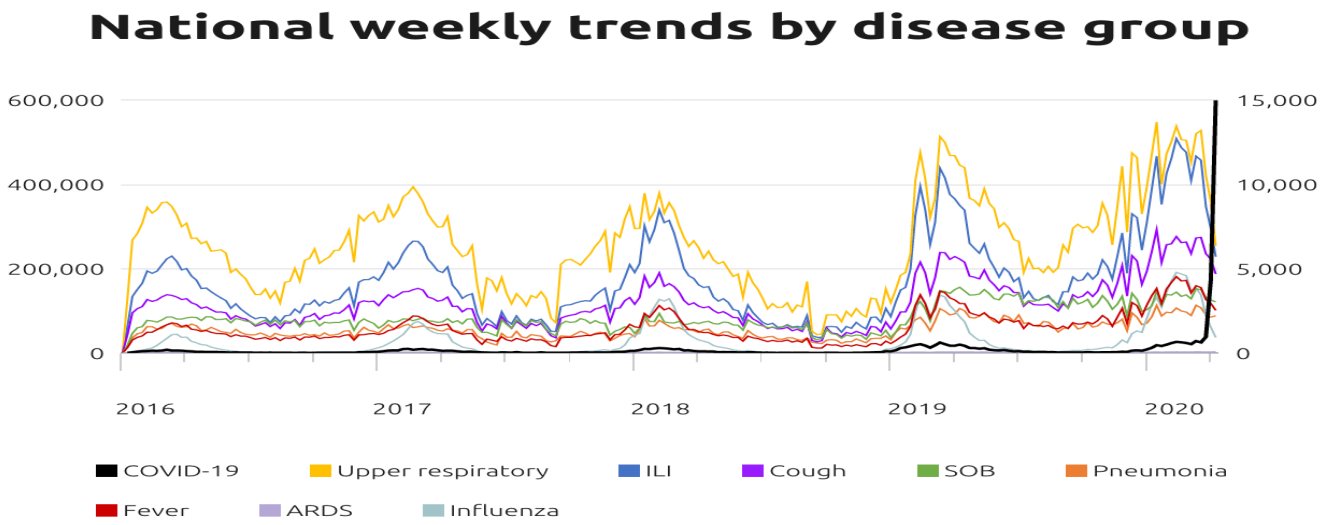
Linkages atypical data

Leveraging Data Aggregators

- Many data aggregators who bring together data across many dimensions of health care
- Some provide an opportunity to capture information not readily available to registries
- These same sorts of data could be used to supplement registries similar to other linkages
  - Including claims
  - Clinical laboratory data
  - Pharmacy data – beyond current linkages

Linkages atypical data: Example of one Data Aggregator

- Health Verity – developing a pilot to assess the longitudinal risk of developing Covid infection subsequent to a positive serologic test.
  - >960,000 patients with at least one serologic Covid 19 test)
  - 1.3 million RNA test in patients
  - Ability to monitor longitudinally linked with enrollment data
- Requires clinical laboratory data – which might also include serologic biomarkers indicative of biochemical recurrence (PSA, CEA, CA 19 etc.)
- Enables longitudinal linkage of these patients through encrypted hashed tokens (secure and privacy preserving)

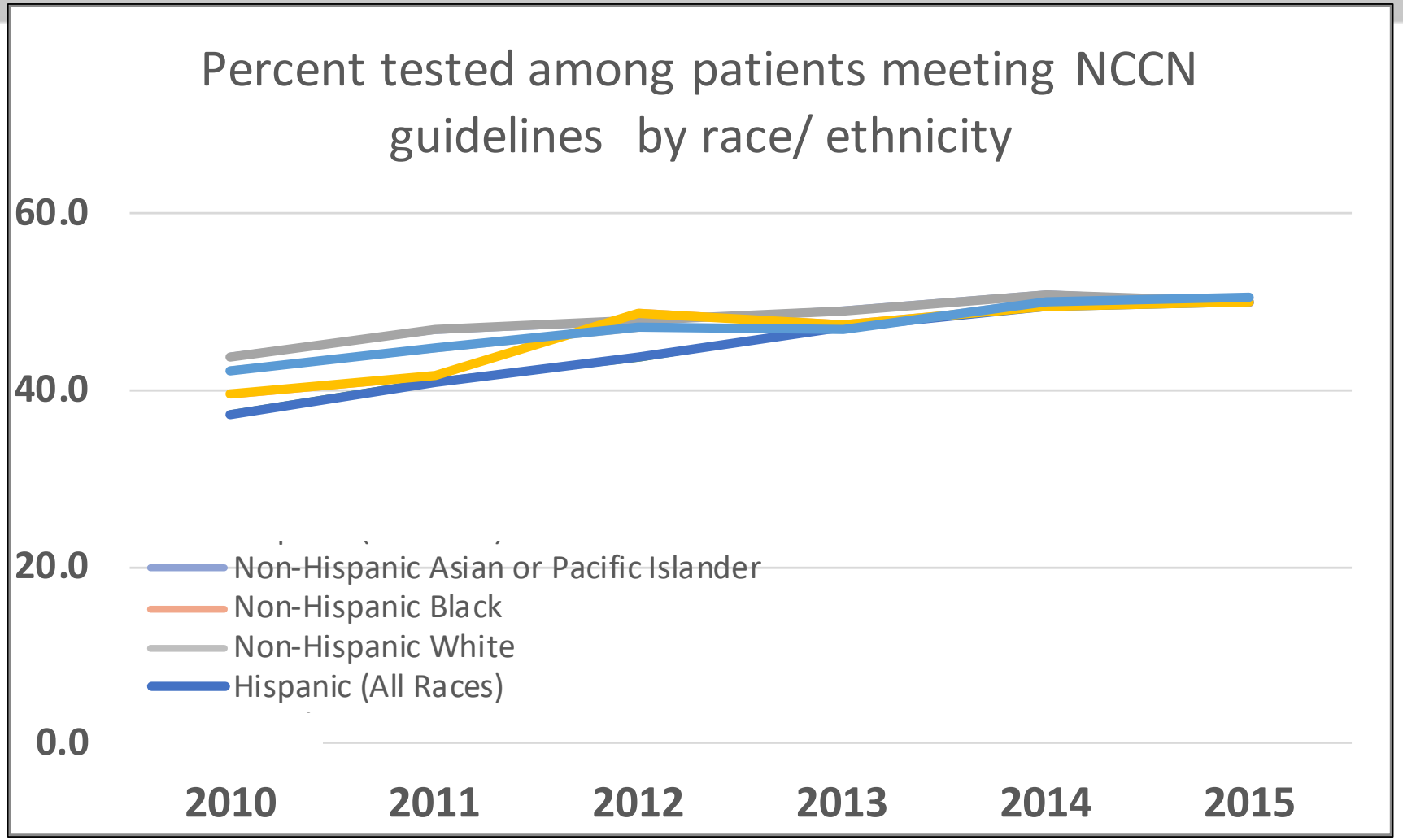


Institute

Linkages for genomic and germline biomarker data

- Genomic pathology labs
  - Biomarkers are
    - Predictive (treatment) and prognostic (outcomes)
    - Required as ***standard of care*** for many cancers
    - Key for monitoring quality of care and understanding differences in outcome
  - It is challenging for registrars to access genomic testing
    - Oncotype DX linkage showed that hospital based reporting missed 43% of test results sent directly to the MD
    - Multigene panels with hundreds of mutations are not feasible to manually report
  - Therefore linking with these companies (as pathology labs) would enable comprehensive capture of data

Example: Evaluating trends in standards of care- disparities in Oncotype DX testing rates



During the initial years (2010-2012), there was some evidence of differential testing by race and ethnicity dependent on age.- recent data suggests disparities are disappearing.

Linkages with genomic data

- SEER now linking with
  - GHI- Oncotype 21 gene, 16 gene and prostate
  - Caris Life Sciences- 71 guideline related biomarkers
  - Castle – Melanoma multi-gene panel
  - Decipher – prostate multi-gene panel
- Developing collaborations to receive key germline mutations that are predictive and prognostic
  - BRCa and other selected actionable mutations
- These linkages couple with key individual biomarkers (SSDI) will provide a rich and complex understanding of each patients cancer and outcomes

Example: Evaluating standards of care- BRCa testing among patients with ovarian (and breast) cancer - CA & GA (2013-2015) *

	Breast Cancer			Ovarian Cancer		
Characteristics	**Total Cases**	**Tested* Cases**	**Proportion Tested* % (95% CI)**	**Total Cases**	**Tested* Cases**	**Proportion Tested* % (95% CI)**
State and year of diagnosis						
California§						
2013	30,367	7,314	24.1 (23.6-24.6)	2,388	707	29.6 (27.8-31.5)
2014	30,012	6,951	23.2 (22.7-23.6)	2,390	732	30.6 (28.8-32.5)
2013-2014	60,379	14,265	23.6 (23.3-24.0)	4,778	1,439	30.1 (28.8-31.4)
Georgia						
2013	8,296	2,066	24.9 (24.0-25.9)	618	206	33.3 (29.6-37.2)
2014	8,410	2,270	27.0 (26.0-28.0)	605	209	34.5 (30.8-38.5)
2013-2014	16,706	4,336	26.0 (25.3-26.6)	1,223	415	33.9 (31.3-36.7)
Race/Ethnicity						
Non-Hispanic (NH) White	48,063	11,635	24.2 (23.8-24.6)	3,701	1,251	33.8 (32.3-35.3)
NH Black	9,039	2,095	23.2 (22.3-24.1)	523	113	21.6 (18.1-25.4)
NH American Indian	207	51	24.6 (18.9-31.1)	19	5	26.3 (9.1-51.2)
NH Asian	9,061	2,034	22.5 (21.5-23.3)	728	229	31.5 (28.1-35.0)
Hispanic	10,715	2,786	26.0 (25.2-26.8)	1,030	256	24.9 (22.2-27.6)

Overall testing (2013-2015) **24% breast cancers and 31% ovarian** cancers.
Substantial variation for ovarian cancer testing ranging from 22% in Black women to 34%in white women
* Kurian et al. JCO April 9, 2019

Automation using NLP and Deep Learning

Automation- NLP extraction from unstructured text (pathology, radiology)

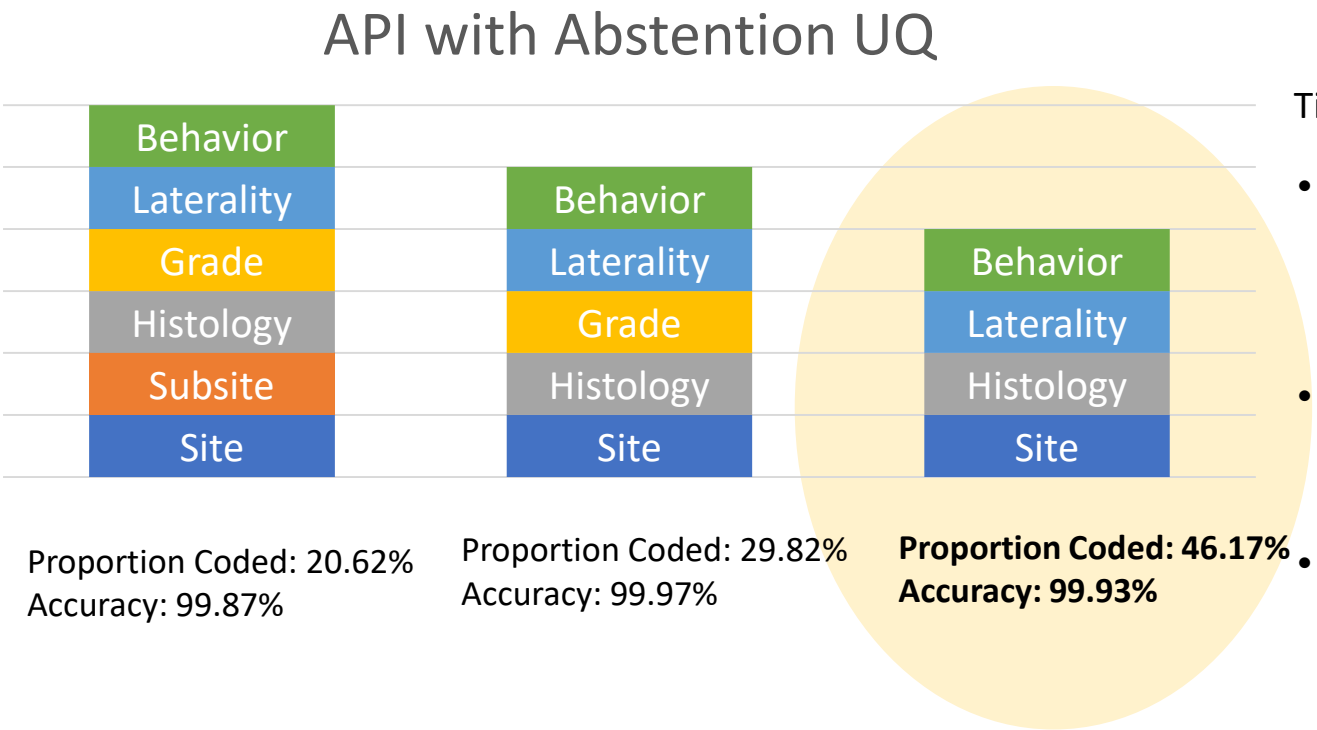
- DOE collaboration to extract:
  - Site, histology, laterality, behavior and grade – testing for use in production
  - Recurrence – testing
    - Initial results 93% accuracy for pathology reports
  - Reportability – refining algorithm
    - Initial results 97% accuracy
  - Targeted biomarkers – initiating this year
    - Existing: ER/PR. HER2Neu, KRAS, MSI
    - New: ALK, EGFR, NRAS, BRAF

Automation- NLP extraction from unstructured text (pathology, radiology)

- The API will be available for use by:
  - Federal partners (CDC, FDA)
  - Registries (central and hospital)
  - Commercial external partners (e.g. CLQ and Tempus

- Leveraging this work to begin automated matching for clinical trials for:
  - Patient eligibility assessment
  - Abstracting structure data from complex unstructured clinical trials eligibility requirements documents
  - Assessing feasibility of a trial accruing to completion

Testing the API for automated extraction across 13 SEER registries and 3,945,946 reports



Time study results:

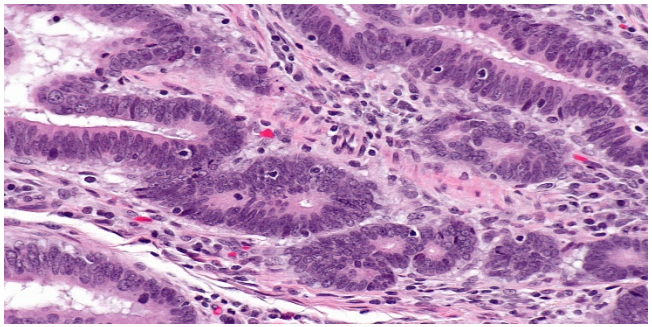
- **Manual** path screening for 5 variables – 1 year, 600,000 path reports: 4,048 hrs (**55sec/report**)
- **AI** path screening on same task: 55 min (**12msec/report**)
- **4500x speed gain to enable near real time cancer surveillance**

- **When API did not abstain, it was > 99.8% correct.**
- **API is ready for workflow implementation to safely reduce workload by nearly 50% for 4 key data elements with nearly perfect accuracy.**

Integrating Novel Technologies Essential for Cancer Surveillance

New technologies

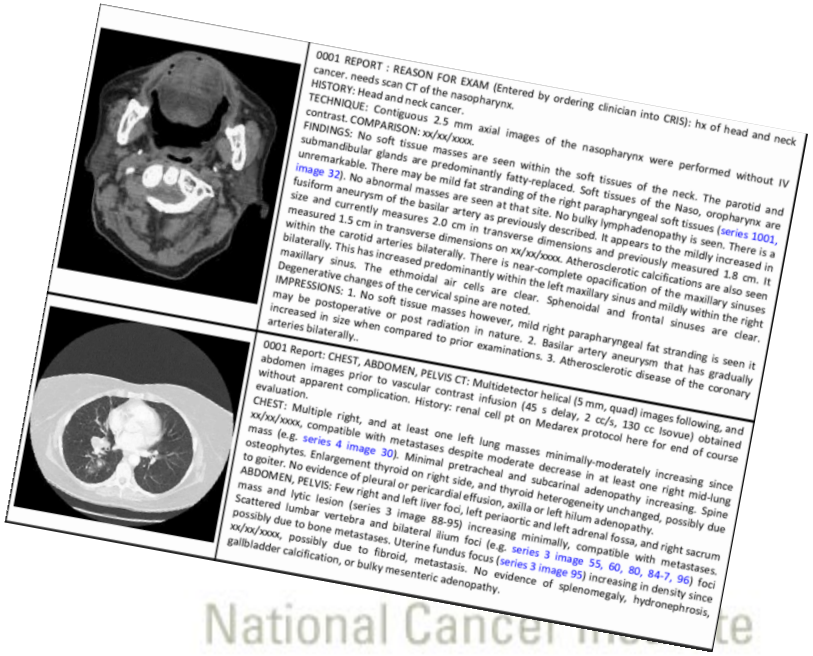
- Digital Pathology Images
  - Provide important information typically missing from pathology report
    - Tumor Infiltrating Lymphocytes – strong prognostic marker
    - Nuclear Pleomorphism



  - We are currently developing a pilot to evaluate a whole slide imaging (WSI) to:
    - assess the ability of registries to capture WSI
    - de-identify and
    - submit for linking with the abstract
  - Ultimately to be made available for researchers and AI experts

New technologies

- Radiology reports and images
  - Important for recurrence, progression of disease
  - May be the only method for diagnosing recurrence
- Beginning work leveraging existing reporting for some SEER registries to access and leverage radiology reports targeting recurrence

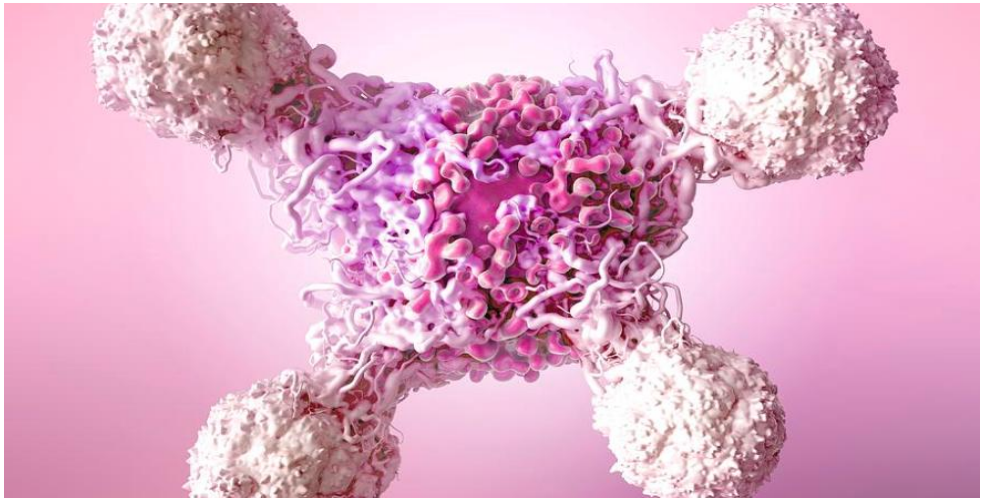


0001 REPORT : REASON FOR EXAM (Entered by ordering clinician into CRIS): hx of head and neck cancer. needs scan CT of the nasopharynx.
HISTORY: Head and neck cancer.
TECHNIQUE: Contiguous 2.5 mm axial images of the nasopharynx were performed without IV contrast. COMPARISON: xx/xx/xxxx.
FINDINGS: No soft tissue masses are seen within the soft tissues of the neck. The parotid and submandibular glands are predominantly fatty-replaced. Soft tissues of the Naso, oropharynx are unremarkable. There may be mild fat stranding of the right parapharyngeal soft tissues (series 1001, image 32). No abnormal masses are seen at that site. No bulky lymphadenopathy is seen. There is a fusiform aneurysm of the basilar artery as previously described. It appears to the mildly increased in size and currently measures 2.0 cm in transverse dimensions and previously measured 1.8 cm. It measured 1.5 cm in transverse dimensions on xx/xx/xxxx. Atherosclerotic calcifications are also seen within the carotid arteries bilaterally. There is near-complete opacification of the maxillary sinuses bilaterally. This has increased predominantly within the left maxillary sinus and mildly within the right maxillary sinus. The ethmoidal air cells are clear. Sphenoidal and frontal sinuses are clear. Degenerative changes of the cervical spine are noted.
IMPRESSIONS: 1. No soft tissue masses however, mild right parapharyngeal fat stranding is seen it may be postoperative or post radiation in nature. 2. Basilar artery aneurysm that has gradually increased in size when compared to prior examinations. 3. Atherosclerotic disease of the coronary arteries bilaterally..

0001 Report: CHEST, ABDOMEN, PELVIS CT: Multidetector helical (5 mm, quad) images following, and abdomen images prior to vascular contrast infusion (45 s delay, 2 cc/s, 130 cc Isovue) obtained without apparent complication. History: renal cell pt on Medarex protocol here for end of course evaluation.
CHEST: Multiple right, and at least one left lung masses minimally-moderately increasing since xx/xx/xxxx, compatible with metastases despite moderate decrease in at least one right mid-lung mass (e.g. series 4 image 30). Minimal pretracheal and subcarinal adenopathy increasing. Spine osteophytes. Enlargement thyroid on right side, and thyroid heterogeneity unchanged, possibly due to goiter. No evidence of pleural or pericardial effusion, axilla or left hilum adenopathy.
ABDOMEN, PELVIS: Few right and left liver foci, left periaortic and left adrenal fossa, and right sacrum mass and lytic lesion (series 3 image 88-95) increasing minimally, compatible with metastases. Scattered lumbar vertebra and bilateral ilium foci (e.g. series 3 image 55, 60, 80, 84-7, 96) foci possibly due to bone metastases. Uterine fundus focus (series 3 image 95) increasing in density since xx/xx/xxxx, possibly due to fibroid, metastasis. No evidence of splenomegaly, hydronephrosis, gallbladder calcification, or bulky mesenteric adenopathy.

New technologies

- Circulating DNA
  - Increasingly used for diagnosing recurrence (liquid biopsies)
  - As we improve this methodology – disease free interval may be continuously shortened



In Summary

- Cancer surveillance must become more agile in the data we collect and methods used
- This will require us to use innovative methods that:
  - Capture data in real time
  - Do not add to the burden of manual data collection
  - Permit the inclusion of complex data sources such as genomic panels or pathology digital images

Thank you

NIH NATIONAL CANCER INSTITUTE