Virtual Pooled Registry Pilot Linkage with Large Cohort Studies

NAACCR Annual Meeting

June 15, 2016

Castine Clerkin, NAACCR

Background

- Multi-state research is costly
- Researchers must link with multiple registries, each with their own IRB review process
- Linkages may yield few or no matches
- Value in linking for case counts prior to IRB application

Virtual Pooled Registry (VPR)

- Recently funded by NCI
- Multi-registry cohort linkages with submission of a single research file
- Linkage performed at the registries
  - Standardized software and protocol
- Release of match case counts, not patient identifiers
- Allow researchers to determine which registries to approach for release of identifiers

Research Study Cohorts for Pilot Linkage

- 2 major pilot linkages with large research cohorts

- Pilot linkages will be used to measure
  - Real world linkage performance with large cohorts
  - Cost savings and efficiency (time required)
  - Completeness of data (compared with traditional process linking registry by registry)
  - Best practices; Value and benefits of VPR

ATSDR Camp Lejeune Cohort (N=340K)

- Marines and civilians from Camp Lejeune and Camp Pendleton (control group)
  - Camp Lejeune (Marines ~ 162K; Civilians ~ 9K)
  - Camp Pendleton (Marines ~ 160K; Civilians ~ 8K)
- Potential exposure to contaminated drinking water
- Addresses located in 2011
  - Only included 50K of the 160K Camp Pendleton marines

Registry Recruitment

- All NPCR and SEER central cancer registries approached
  - Regional SEER registries participated if state unable
  - 43 state registries, District of Columbia, and the Seattle SEER registry participated
    - 5 state registries unable due to staffing, resources, and competing priorities
    - 2 state registries required IRB contact for linkage and release of match counts
    - 1 state registry could not get software approved in time

NCI Radiation Technologists Cohort (N=146K)

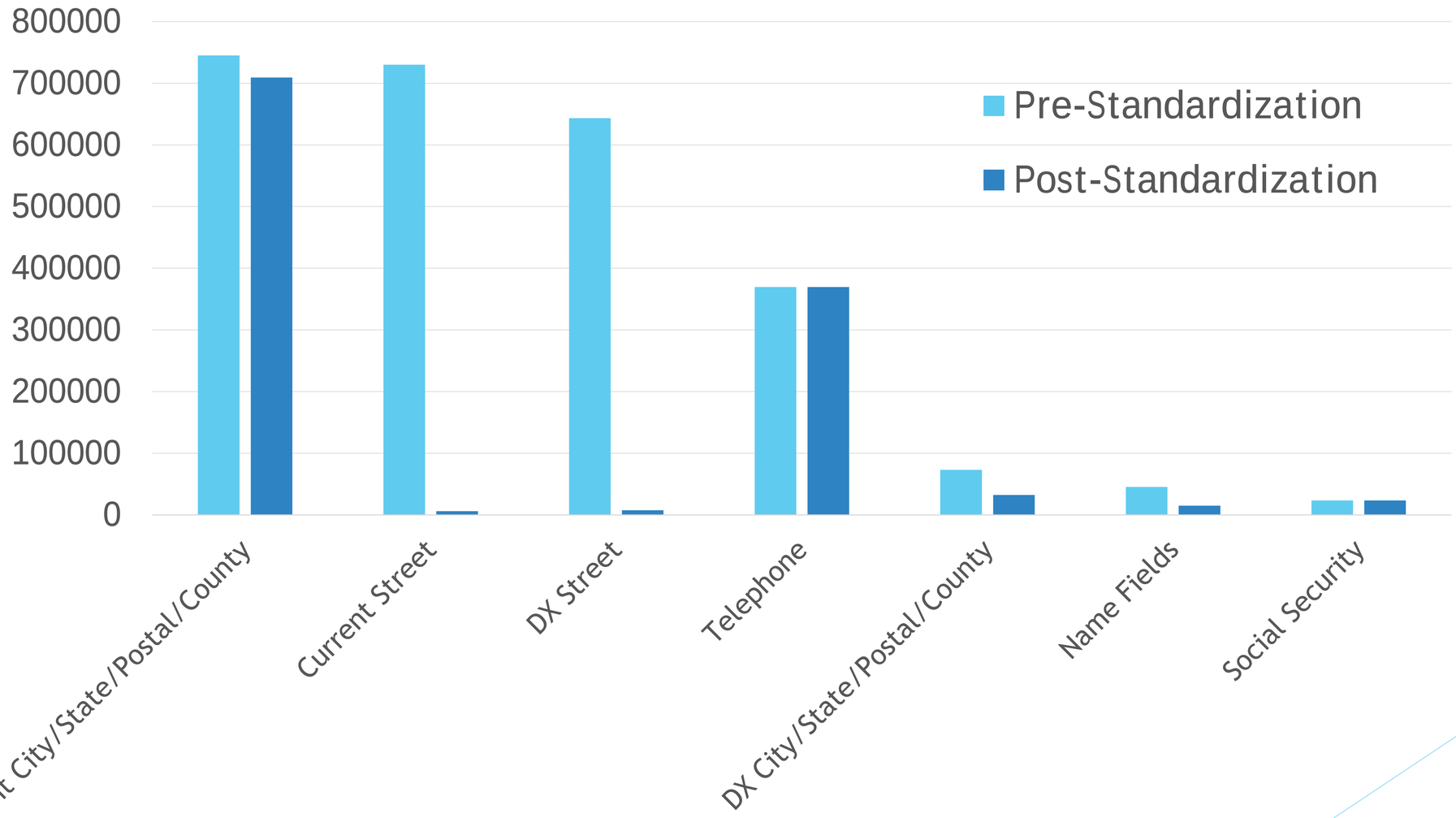
- Cohort of U.S. Radiation Technologists
  - Exposure estimates
- Surveyed every 10 years since 1980
  - Resource-intensive medical record validation of cancer
  - Declining response rates
- Linkage with registries is new initiative
  - Amendment to existing protocol with NCI Special Studies IRB
- VPR cohort file linkage: Early July 2016
  - Anticipate linking with same 45 registries

Linkage Software

- Link Prep: Edits and standardizes the data
  - Run by both researcher and registry
  - Researcher resolved edits, registries did not
- BigMatch: US Census software that efficiently identifies matched pairs
- BMReader: Evaluates the quality of the matched pairs and assigns to high quality, manual review, or non-match.
  - Produces count of high quality and manual review matches by diagnosis year

RESULTS

Edits and Standardization



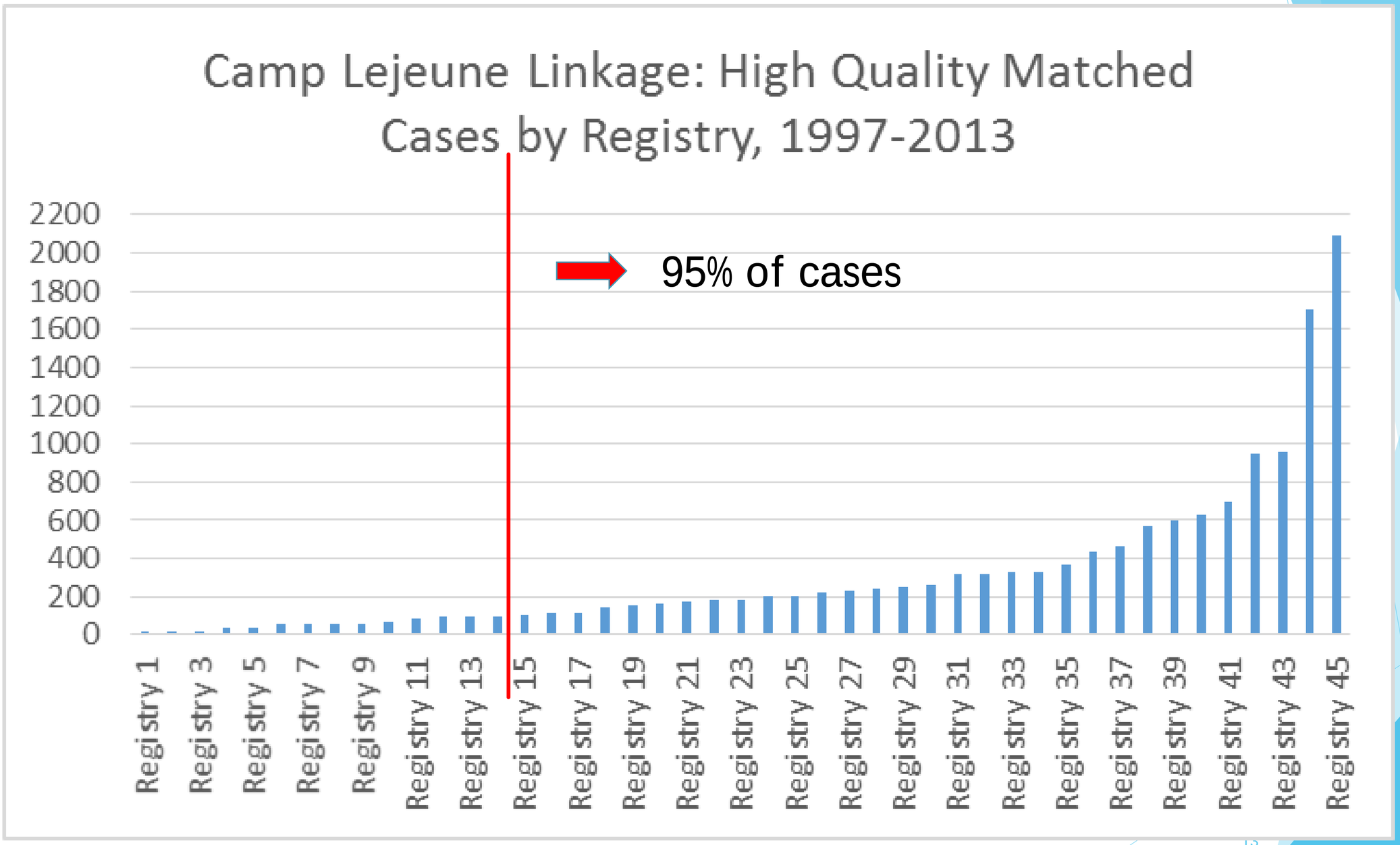
Time and Cost

	Software Runtime (minutes)	Total Staff Time (hours)	Actual Staff Cost	Routine Charges
Range	0.8 – 82.0	1.5 – 21.0	$71 - $2,185	$0 – $17,488
Mean	15.1	7.43	$665	$2,309
Median	8.8	6.25	$567	$325

Total Staff Cost: $26,600
Total Routine Charges: ~$100,000

High Quality (true) Match Results

- Registries included varying years in linkage
  - 1997-2013 were common across all registries
- All Years: 17,412 cases and 16,160 patients
- 1997-2013: 14,514 cases and 14,213 patients


Camp Lejeune Linkage: High Quality Matched Cases by Registry, 1997-2013
95% of cases
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Registry 1
Registry 3
Registry 5
Registry 7
Registry 9
Registry 11
Registry 13
Registry 15
Registry 17
Registry 19
Registry 21
Registry 23
Registry 25
Registry 27
Registry 29
Registry 31
Registry 33
Registry 35
Registry 37
Registry 39
Registry 41
Registry 43
Registry 45

Manual Review Match Results

- Manual Review Matches
  - 63,938 cases and 57,573 patients

	Count	Percent
Lejeune civilians	2291	4.0%
Lejeune marines	13057	22.5%
Pendleton civilians	1967	3.4%
Pendleton marines	40637	70.1%

- Voluntary Manual Review
  - 15 central registries, 12 reviewed all matches

Manual Review Results

- False Positives among High Quality
  - 6/4,178 = 0.14%
- True Matches among Potentials
  - 158/23,810 = 0.67%
  - Overall 2.4% increase in true matches

- Workgroup established to review results
  - Specifics on which data items did or did not match
  - Modifications to match criteria
  - Reduce volume of manual review matches

Modifications to Linkage Software

- Modification to blocking strategy
- Incorporation of Soundex to facilitate blocking
- Matching and weighting by the 3 sections of SSN and DOB, rather than all characters
- Use of middle name instead of address
- Evaluation of swaps between first name and middle name

Conclusions

- Streamlined, efficient multi-registry linkages are possible
- Standardized software improves data quality and comparability of linkage files
- Staff time/costs were reasonable, but will be reduced with extraction of single linkage file, experience with process, and software automation
- Match counts aid researchers in determining which registries to approach for full IRB application
- Pilot tests provide valuable information to improve linkage process

Thank you!

- All 45 AMAZING registries who participated!
- IMS, Inc.: Andy Lake and Will Howe
- NCI: Lynne Penberthy
- LA SEER Registry: Dennis Deapen
- NAACCR: Betsy Kohler and Recinda Sherman