Using NPI and ePath Facilities to Identify Clinics That Treat or Diagnose Cancer

by Christian Klaus, North Carolina Central Cancer Registry

Outline

- Background
- Study Objectives
- Recent Developments
- Our Approach
- Facilities to be Linked
- Results
- Conclusions
- Related Topic – Database Design to Capture Ownership, Abstraction, Preferred Followback, Name/Address aliases, CRM

Background

- Geospatial analysis of cancer incidence is impacted by missing caseload whose distribution across space and facilities is typically unknown. The amount has been estimated with capture-recapture methods and linkage to Medicare claims data[1,2].
- [1]McLish and Penberthy, 1997 'Ability of Medicare Claims Data and Cancer Registries to Identify Cancer Cases and Treatment', American Journal of Epidemiology 145:3, pp 227-233.
- [2]McLish and Penberthy, 2004: 'Using Medicare Data to Estimate the Number of Cases Missed by a Cancer Registry', Medical Care 42:11, pp 1111-1116.

Study Objectives

- We would like to be able to assign count of ePath reports (consolidated to patient) to treaters and/or diagnosers that are not reporting or being reported for.
- This is to enable missing case estimates corresponding to consolidated or unconsolidated ePath reports, at county and sub county scales.
- Project focus is to evaluate the strategy of estimating missing cases by first finding treating and/or diagnosing facilities that are not reporting, or being reported for, by merging ePath and NPI facilities with NC CCR facilities.
- Part 1: match all ePath facilities against all NC CCR facilities and NPI to understand match rate;
- Part 2: identify ePath ordering facilities for 13959 path reports not matched to NC CCR existing cases, and categorize them as reporters or non-reporters.

Recent Developments

- Recent communications from the COC indicate that CCRs are responsible for collecting and maintaining ownership and abstracting relationship data for facilities that treat or diagnose cancer, if the facilities themselves can't provide these data.

Our Approach

- We wanted to use a process of elimination to find the non-reporters in ePath and NPI. The prerequisite to the process of elimination is to attempt to capture from different sources, and categorize all treaters and diagnosers within a state into exclusive categories (controlling for name/address aliases as best we could):

1) Facility pairs that mutually recognize ownership relationship between them (owner/owned).
2) Facility pairs that mutually recognize abstraction relationship between them exclusive of ownership relationship (abstractor/abstractee).
3) Facilities claiming an ownership or abstraction relationship.
4) Facilities whose abstraction relationships can't be resolved with facility attributes, because responsibility for their cases has to be decided at case level.
5) Known Reporters

Facilities that treat and diagnose, and are not in the above groups, are likely non-reporters.

Abstraction Relationships at Case Level

- There is currently no NAACCR data item that captures the reporting facility ID of an independent clinic (abstractee) whose case was abstracted by another facility (abstractor).
- These cases correspond to class of case 42.
- These cases are currently credited to the abstracting facility (abstractor).
- We have proposed a new data item – 'Abstracted on Behalf Of', to capture the abstractee facility ID.

Facilities to Be Linked: NC CCR Reporting Facilities

- We asked NC hospitals or hospital networks to give us lists of clinics that they own or abstract on behalf of. Most but not all were in a position to share those data. We received data on 1170 clinics, of which :
- 772 were owned, 750 were owned with all cases abstracted, 22 were owned with some cases abstracted,
- 62 were abstracted for, 41 were abstracted for, for all cases, 21 were abstracted for, for some cases.
- NC has 237 current reporting sources, and 252 historic. Of the current, 87 are hospitals that provide bulk of NC CCR cases.
- 46 (current) hospital networks, which are affiliated with 1170 clinics.

Facilities to Be Linked: Other Data Sources Used

- We obtained hospital level ownership data (annual licensing and renewal) from NC Division of Health Service Regulation.
- We obtained NPI Type 2 records (facilities) from CMMS current to December 2014.
- ePath facilities derived from ~100K Emarc_plus database records (path reports) current to January 2015. Approximately1454 path ordering facilities.
- ePath record fields that we used were patient demographic data and ordering facility attributes (name, address, ZIP code) from ORC, OBR and PID tables.

ePath Facility Primary Key Issues

- The vast majority of our ePath reports do not have a solid (NPI or other PK) ordering facility primary key.
- For some ePath reports, the ordering physician is listed as the facility; sometimes the physician name=facility name, other times the physician just works at the facility.
- Reports come in with facility name and address aliases and abbreviations.
- For some ePath records the facility name and/or address is incorrect.
- Even when an NPI is submitted, for many facilities more than one NPI exists in NPI data.
- We needed to assign an initial PK, then create PK/NPI (one to many, many to many) lookup table afterward.

ePath Facility Primary Key Strategy

- To assign initial primary key, group ePath facilities by facility name, address and ZIP code (assume these as PK components).

- 1) use an auto number field to assign initial PK
- 2) use geocoding to compare addresses, group by X and Y to control for address aliases. Identify name aliases by running find duplicates on X and Y. This will identify name aliases that a name sort would not find. Be aware that what look like name aliases may be separate NPIs (example: clinic function name, clinic business/billing name).
- 3) Use SAS text string similarity functions to compare facility names, group by similarity to control for name aliases, where PKs are different. This will help identify incorrect addresses.
- 4) Selectively reassign PKs (one to many) for duplicates, to control for more than one facility at same address.
- Understood: initial PK will not detect all name or address aliases, so a small amount of duplicate facilities will remain.

ePath Facility Primary Key Strategy, con't

- After assignment of initial PK, we want to create an ePath PK/NPI lookup table, to capture one to many or many to many ePath PK/NPI relationships.
- Record level, time consuming process
- Benefit 1: Capture reliable facility name and address aliases for future batch linkage, and identify aliases that initial PK did not identify
- Benefit 2: Enables potential to leverage CMMS work to manage facility turnover by monitoring NPI by comparison of periodic NPI updates

ePath Value Add, con't

- Of 13959 ePath reports that did not link to NC CCR cases, 4403 did not get assigned an ePath report number. 4172 of these were successfully linked in Emarc Plus database to OBR table via PID table, using patient first, last, and date of birth (exact match).

Facility Linkage Considerations

- Facility name/address linkage is generally not as good as patient (names, DOB, SSN, other attributes) linkage.

Part 1: Link ePath Facilities to NC CCR Facilities

- Linkage on: name, address, name alias, address alias and 5 digit ZIP code using SAS 9.4 string similarity function
- 1170 owned/abstracted for clinics vs. 273 NC CCR current facilities and clinics: Overlap of 23 records.
- 1454 (plus 293 name and/or address aliases) ePath Facilities (Jan 2015) vs. 1420 NC CCR current facilities and clinics:
- Epath Facilities Linked: 207 if aliases consolidated (14.2%)/271 if aliases unconsolidated (15.5%).

Part 1: Link ePath Facilities to NPI Type 2 Records (Facilities)

- Linkage on: ePath Facility name, name alias, address, address alias and 5 digit ZIP code vs. 4 NPI fields
- NPI- 3 interactive reviews, one each for N005-Orgname, N012-Oorgname, N029-Physical Address, N033-Physical Address Postal Code
- 1454 (plus 293 name and/or address aliases) ePath Facilities (Jan 2015) vs. Dec 2014 NPI Type 2 (Facilities) data: 1054 (aliases unconsolidated), or 60.3%

Link to NPI, con't

- NPI against 1420 NC CCR current facilities and clinics:
- 495 out of 1420 (34.8%)
- Joined linked tables on NPI to increase ePath Facility/NC CCR linkage by 20 facilities (15.5% to 16.6%).
- A better linkage can be had between ePath Facilities and NPI Facilities against CCR Facilities than against CCR Facilities alone, on account of 2 name fields in NPI.

Part 2 Results

- From 13959 ePath reports that did not match to an existing NC CCR case, we identified 64 facilities that ordered the reports, that did not appear to be owned or abstracted for by other facilities, based on our facility linkages.

Conclusions

- Linkage between medical facilities is generally weaker than linkage between patients, on account of aliases and fewer linkage fields.
- Identifying non reporters through ePath is a challenge, if a facility primary key is lacking, and because not all aliases, owned/abstracted for clinics may be reported to CCR
- Large number of matches between ePath facilities and NPI not made because of different house numbers.
- NPI linkage is more accommodating to organizational hierarchies, with 2 name fields, that enable matching at same organizational level.
- Path labs are oftentimes using facility names directly from NPI.
- NPI linkage to ePath facilities is very helpful in that its higher match rate makes it easier to differentiate name and address aliases from abbreviations or incorrect data, as compared to just linking ePath facilities to CCR reporters and their affiliates.

Facility/Physician Centric Database Design

- With facility and physician centric database tables in CCR database, the task of managing:
- 1) facility name and address aliases
- 2) facility currency
- 3) facility affiliations (preferred followback, ownership, abstraction relationships)
- 4) facility-CCR relationship attributes(and might as well include customer relationship management functionality)
- 5) physicians, physician currency, physician name/address aliases and physician affiliations with facilities
- 6) facility/NPI lookup tables
- 7) allows for uncertainty among facilities and facility relationships until contact can be made
- …..can be spread among all staff in a CCR instead of one or more designated people.

Questions?

- Christian.Klaus@dhhs.nc.gov