Topic 1

IMM/Scrub:
Unduplication of Vaccination History Records in Childhood Immunization Registries (completed)

Overview

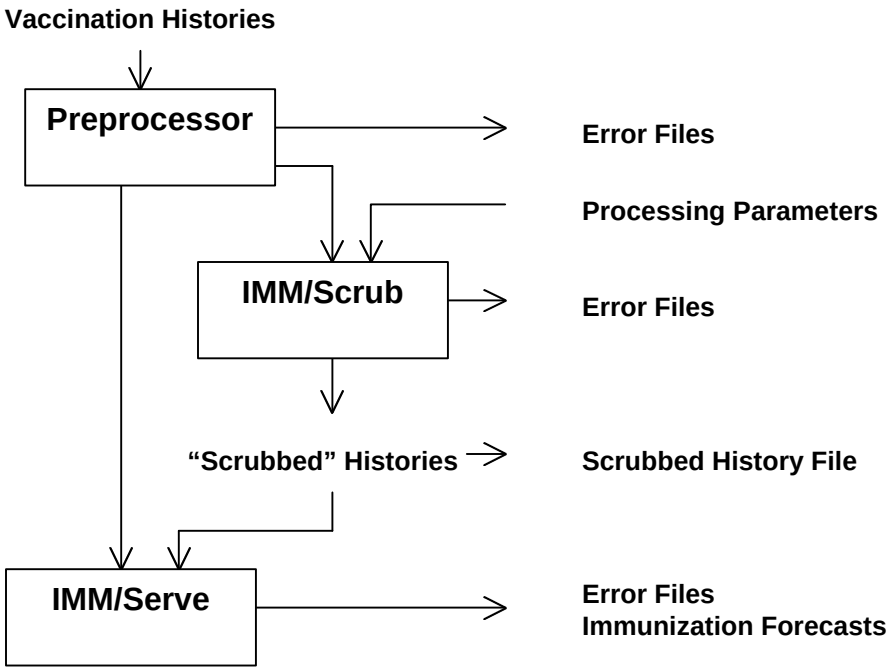
- **IMM/Scrub is a pilot tool for unduplication of vaccination history records in immunization registries.**

- **A great deal of attention has been paid to unduplicating immunization records at the demographic level.**

- **No general tools are available, to our knowledge, for vaccination history duplicates.**

- **This talk describes IMM/Scrub and our experience with three immunization registries:**

  - **a state registry (over 430,000 patients),**
  - **a metropolitan area registry (over 180,000 patients),**
  - **a medical center clinic (roughly 7,500 patients).**



Overview of IMM/Scrub's Operation

- **IMM/Scrub currently operates in "batch" mode.**
- **It takes as input a file of immunization histories and a set of user-defined processing parameters.**
- **It outputs several files for manual inspection and further computer-based analysis.**
- **A future project could build a graphical-user-interface to let the user inspect the output files interactively.**

Complicating Factors in Unduplicating Vaccination Doses

- **Some doses may have dose numbers and others may not.**

- **Two doses may have different dose numbers.**

- **Doses may specify different preparations within a series,**

  - e.g., DT, DTP, DTaP, and Td.

- **One dose may indicate a combination vaccine (DTP-Hib) and the other dose may indicate one of those vaccines (Hib).**

- **Two doses may have slightly different dates.**

- **Several of these problems may apply at the same time.**

Dealing with these Complexities: Three Examples

- **Certain duplications can be corrected in a completely automated fashion.**

  e.g., - same vaccine (same HL7 code), and
  - same date, and
  - same dose number (or no dose number).

- **It may be clear that the problem can never be corrected automatically, and that human judgment will be required.**

  e.g., - one dose specifies a combination (Hib-DTP), and
  - the second dose specifies DTaP, and
  - the two doses have a different dose number, and
  - the two doses have dates that differ by one day.

- **Registry staff may be comfortable letting certain duplications be corrected automatically if they understand how the error occurred.**

  e.g., Two doses are 2-3 days apart but otherwise identical, and come from a provider that often enters doses both with they are administered and a few days later when they are billed.

Ten Duplication Tests
Currently Handled by IMM/Scrub

- **Identical Vaccines on the Same Date**
- **Subsumed Doses on the Same Date**
- **Coincident Doses on the Same Date**
- **Identical Dose Numbers in a Vaccine Series**

- **Identical Vaccines within a Date Window**
- **Subsumed Doses within a Date Window**
- **Coincident Doses within a Date Window**

- **Identical Vaccines within a Fixed Day Neighborhood**
- **Subsumed Doses within a Fixed Day Neighborhood**
- **Coincident Doses within a Fixed Day Neighborhood**

Duplicate Test:
Identical Vaccines on the Same Date

- **Here both doses specify the same vaccine (same HL7 code) and the same date.**
  - **Both may have no dose number.**
  - **Both may have the same dose number.**
  - **One dose may have a dose number and the other may not.**
  - **Dose numbers may disagree.**

- **When IMM/Scrub is run, user-defined processing parameters indicate for each duplicate test:**
  - **whether that test should be performed,**
  - **how each set of conditions should be handled.**

- **For this test, a reasonable strategy may be:**
  - **No dose numbers: eliminate either**
  - **Same dose number: eliminate either**
  - **Only one dose number: eliminate unnumbered dose**
  - **Different dose numbers: bypass dose elimination**

Duplicate Test:
Subsumed Doses on the Same Date

- **One dose indicates a combined vaccine (e.g., DTP-Hib); the other indicates one of the components (e.g., Hib).**

- **For this test, a reasonable strategy may be:**

  - **No dose numbers: eliminate single dose**
  - **Same dose numbers: eliminate single dose**
  - **Different dose numbers: bypass dose elimination**
  - **One dose number (combo): eliminate single dose**
  - **One dose number (single): synthesize new dose record?**

- **IMM/Scrub is currently only able to eliminate doses, not synthesize new dose records.**

- **An additional issue is that most immunization registries allow each vaccine dose to have at most a single dose number. (But HepB-Hib may be HepB dose 2 and Hib dose 1.)**

Duplicate Test:
"Coincident" Doses on the Same Date

- **Two doses are in the same series but indicate different vaccines, e.g., DT, DTP, DTaP, or Td.**

- **Using the processing parameters, the user may specify a preference order among the HL7 codes for each series.**

- **For this test, a reasonable strategy may be:**

  - **No dose numbers:** **eliminate less preferred**
  - **Same dose numbers:** **eliminate less preferred**
  - **Different dose numbers:** **bypass dose elimination**
  - **Only one dose number:** **bypass dose elimination (or synthesize new record)**

Duplicate Test:
Identical Dose Numbers in a Vaccine Series

- **Dose pairs with the same dose number and different dates.**

	Duplicate doses detected (conservative | liberal)	Duplicate doses eliminated (conservative | liberal)
State Registry (431,024 patients) (prior to extensive unduplication)		
identical doses, same date	**71,849 | 71,848**	**70,823 (99%) | 71,848 (100%)**
subsumed doses, same date	**2,464 | 2,464**	**688 (28%) | 693 (28%)**
coincident doses, same date	**5,947 | 5,933**	**5,228 (88%) | 5,610 (95%)**
duplicate dose numbers	**16,600 | 15,978**	**0 | 15,978 (100%)**
total duplicate doses above	**96,860 | 96,223**	**76,739 (79%) | 94,129 (98%)**
patients involved	**46,224**	**35,299 (76%) | 46,213 (~100%)**
% of patients in database	**11%**	**8% | 11%**
City Registry (186,661 patients) (no dose numbers, operational for several years)		
identical doses, same date	**1,653**	**1,653 (100%)**
subsumed doses, same date	**1,931**	**1,931 (100%)**
coincident doses, same date	**3,345**	**3,345 (100%)**
duplicate dose numbers	**NA**	**NA**
total duplicate doses above	**6,929**	**6,629 (100%)**
patients involved	**3,596**	**3,596 (100%)**
% of patients in database	**2%**	**2%**
Medical Center Database (7,479 patients) (operational for two years)		
identical doses, same date	**22**	**20 (91%) | 22 (100%)**
subsumed doses, same date	**28**	**14 (50%) | 20 (71%)**
coincident doses, same date	**31**	**5 (16%) | 5 (16%)**
duplicate dose numbers	**225 | 224**	**0 | 224 (100%)**
total duplicate doses above	**306 | 305**	**39 (13%) | 271 (89%)**
patients involved	**234**	**28 (12%) | 207 (88%)**
% of patients in database	**3%**	**0.3% | 3%**

Duplicate Tests:
Identical Vaccines within a Date Window
Subsumed Doses within a Date Window
Coincident Doses within a Date Window

- **These three tests are similar to first three discussed, but involve a different date.**

- **Using the processing parameters, the user sets a "date window" (DW) of 30 days or less.**

- **For example, if DW = 7, all the duplicate doses identified will differ by 1-7 days.**

- **The date window extends across the beginning and end of a month.**

	Duplicate doses detected (conservative | liberal)	Duplicate doses eliminated (conservative | liberal)
State Registry (431,024 patients)		
identical doses, DW = 3	**1,760 | 1,757**	**0 | 1,757 (100%)**
identical doses, DW = 7	**2,793 | 2,786**	**0 | 2,786 (100%)**
identical doses, DW = 14	**4,402 | 4,386**	**0 | 4,386 (100%)**
identical doses, DW = 21	**6,213 | 6,180**	**0 | 6,180 (100%)**
subsumed doses, DW = 3	**50**	**7 (14%)**
subsumed doses, DW = 7	**63**	**12 (19%)**
subsumed doses, DW = 14	**78**	**18 (23%)**
subsumed doses, DW = 21	**94**	**30 (32%)**
coincident doses, DW = 3	**79**	**0 | 79 (100%)**
coincident doses, DW = 7	**118 | 117**	**0 | 117 (100%)**
coincident doses, DW = 14	**170 | 169**	**0 | 169 (100%)**
coincident doses, DW = 21	**212 | 211**	**0 | 211 (100%)**
City Registry (186,661 patients)		
identical doses, DW = 3	**6,219 | 6,196**	**0 | 6,196 (100%)**
identical doses, DW = 7	**9,108 | 9,029**	**0 | 9,029 (100%)**
identical doses, DW = 14	**13,289 |13,110**	**0 | 13,110 (100%)**
identical doses, DW = 21	**15,545 |15,267**	**0 | 15,267 (100%)**
subsumed doses, DW = 3	**874 | 871**	**874 (100%) | 871 (100%)**
subsumed doses, DW = 7	**1,244 | 1,238**	**1,244 (100%) | 1,238 (100%)**
subsumed doses, DW = 14	**1,842 | 1,822**	**1,842 (100%) | 1,822 (100%)**
subsumed doses, DW = 21	**2,111 | 2,071**	**2,111 (100%) | 2,071 (100%)**
coincident doses, DW = 3	**1,421 | 1,420**	**0 | 1,420 (100%)**
coincident doses, DW = 7	**1,808 | 1,807**	**0 | 1,807 (100%)**
coincident doses, DW = 14	**2,405 | 2,400**	**0 | 2,400 (100%)**
coincident doses, DW = 21	**2,622 | 2,601**	**0 | 2,601 (100%)**
Medical Center Database (7,479 patients)		
identical doses, DW = 3	**0**	
identical doses, DW = 7	**2**	**0 | 2 (100%)**
identical doses, DW = 14	**5**	**0 | 5 (100%)**
identical doses, DW = 21	**12**	**0 | 12 (100%)**
subsumed doses, DW = 3, 7, 14, 21	**0**	
coincident doses, DW = 3, 7, 14, 21	**0**	

Additional Problems Identified by IMM/Serve

- **The user may request that the IMM/Serve forecasting program be invoked.**

- **IMM/Serve is called twice for each case, once with the original history and once with the scrubbed history.**

- **IMM/Serve identifies a variety of errors.**

	State	City	Med Ctr
Dose screened (given too early)	**45,459**	**24,020**	**1,172**
Dose number too big for series	**18,043**	**-**	**111**
Too many doses prior to a numbered dose	**9**	**-**	**0**
Numbered doses not in chronological order	**9,645**	**-**	**14**

- **IMM/Serve also indicates whether the data was "clean" enough to allow a successful forecast for each series.**

- **IMM/Serve's forecast may be inspected by the user to see if it looks reasonable, or whether it suggests the presence of further errors.**

Current Status

- **IMM/Scrub is a pilot implementation:**
  - **to assess the level and types of duplicates in several immunization registries,**
  - **to explore the design issues involved in detecting and correcting those duplicates.**