Topic 2:

Exploring Three Approaches for Handling Incomplete Patient Histories in a Computer-based Guideline for Childhood Immunization (completed)

Overview

- **Data quality is an important concern in any clinical database.**

- **Immunization registries have several types of data quality problems, including:**

    - data errors
    - duplicate records at the demographic level
    - duplicate records at the vaccination level
    - incomplete vaccination histories (missing doses)

- **Incomplete vaccination histories (containing missing doses) typically prevent the automated generation of immunization recommendations.**

Two Questions

- **How prevalent are incomplete histories in real immunization registries?**

- **To what extent can an immunization forecasting program accommodate missing doses?**

- **We explored these questions using IMM/Serve, a forecasting program for childhood immunization.**

IMM/Serve:
An Immunization Forecasting Program

• takes as input a child's immunization history

• produces immunization recommendations:
    • which vaccine doses are due now
    • which vaccine doses should be scheduled next

IMM/Serve: Example Input

Date today: **11/1/96**

Date of birth: **8/1/95**

HepB: **8/1/95 [1], 10/2/95 [2], 2/4/96 [3]**

DTP: **10/2/95 [1], 12/4/95 [2], 2/4/96 [3]**

Hib: **PRPOMP 10/2/95 [1], PRPOMP 12/4/95 [2]**

OPV: **10/2/95 [1], 12/4/95 [2]**

MMR: **8/3/96 [1]**

IMM/Serve: Example Output

The following immunization(s) are due on 11/1/1996:
    DTaP 4
    Hib 3
    OPV 3
    VAR 1

The following immunization(s) will be due:
    D/T Series dose 5, on or after 8/1/1999 but before 8/1/2001.
        (if DTaP 4 is given on 11/1/1996)
    OPV 4, on or after 8/1/1999 but before 8/1/2001.
        (if OPV 3 is given on 11/1/1996)
    MMR 2 or Me 2, on or after 8/1/2000.

The following vaccine series are either complete or no longer relevant for this case:
    HepB

Advisory Notices:
    In Oregon, the D/T Series dose 5 and the OPV/IPV Series dose 4 are due prior to kindergarten entry.

Examples Vaccination Histories

Complete Histories:

    DTP: 1/1/97, 3/1/97, 5/1/97

    DTP: 1/1/97 [1], 3/1/97 [2], 5/1/97 [3]

    DTP: 1/1/97, 3/1/97 [2], 5/1/97 [3]

Incomplete Histories:

    DTP: 1/1/97 [1],   5/1/97 [3]

    DTP: 5/1/97 [3]

    DTP: 6/1/97,   5/1/98 [4]

    DTP: 1/1/97 [1], 3/1/97 [2], 2/1/98

    DTP: 1/1/97 [1], 3/1/97 [2], ?/?/? [3]

Adapting IMM/Serve
to Handle Incomplete Histories

- **Design philosophy**

  - **Make as few changes as possible to the existing knowledge base (KB).**

  - **To the extent possible, only modify the underlying inference engine.**

- **Overall approach**

  - **"Manufacture" computer-generated doses that allow the KB logic to operate without modification. (Use earliest date possible.)**

  - **Adapt the inference engine by disabling tests for minimum age & interval for the computer-generated doses.**

Alternative Strategies
for Handling Incomplete Histories

- **Default strategy:** Ignore dose numbers.

- **Mandatory dose number strategy:** Insist that all dose numbers be present.

- **Low dose number strategy:** Lowest numbered dose is treated as correct.

- **Designated dose strategy:** User designates a specific dose number as correct.

Preliminary Assessment of the Approach

- **We used the "low dose number" strategy with a State registry database containing 431,024 patient records.**

- **We first used IMM/Scrub to eliminate most duplicate doses.**

Results

- **60,336 cases (14%) had one or more incomplete series history.**

Incomplete Series Histories

DTP	32,659	7.6%
Hepatitis A	34	0.008%
Hepatitis B	17,910	4.2%
Hib	12,409	2.9%
MMR	5,765	1.3%
Polio	28,929	6.7%
Varicella	38	0.009%
	97,744	

- **These data suggest that:**

  - **Incomplete histories pose a significant problem.**

  - **An immunization forecasting program that accommodates such histories could be very helpful.**

Results (continued)

- **82% (80,437) of the vaccine series with incomplete histories were able to be run by IMM/Serve.**

Successful IMM/Serve Runs

DTP	**28,065**	**86%**
Hepatitis A	**22**	**65%**
Hepatitis B	**15,873**	**89%**
Hib	**9,551**	**77%**
MMR	**5,631**	**98%**
Polio	**21,274**	**74%**
Varicella	**21**	**55%**

- **16,822 failures (97%) were due to data quality problems:**
  - **dose number too high**
  - **dose numbers not in chronological order**
  - **other**

- **For 430 series (2%), there was an insufficient time interval for the missing dose(s).**

- **For 55 series (0.3%), IMM/Serve had to screen a real dose as invalid after inserting its manufactured dose(s) into the series history. In this circumstance, IMM/Serve aborts its analysis of the series.**

Discussion: Lessons Learned

- **The need for different strategies.**

- **Handling incomplete histories by modifying the inference engine vs. augmenting the KB.**

- **Placing limits on the allowed degree of incompleteness.**

- **Data cleaning vs. immunization forecasting: where to draw the line?**

Summary

- **There are many types of data quality problems in clinical databases.**

- **Incomplete vaccination histories are common in immunization registries.**

- **A computer-based immunization forecasting program can be designed to accommodate much of this incompleteness.**

- **Different strategies may be used.**