Exploring Three Approaches for Handling Incomplete Patient Histories in a Computer-Based Guideline for Childhood Immunization

Perry L. Miller, M.D., Ph.D. [1], Sandra J. Frawley, Ph.D.[1,2], Frederick G. Sayward, Ph.D.[1,2]
[1]Center for Medical Informatics, Yale University School of Medicine, New Haven, CT
[2]Medical Decision Associates, Inc., New Haven, CT

A significant problem faced by immunization registries is that the dates of a patient's previous vaccinations may not be known. These incomplete histories can pose a problem when attempting to use a computer-based guideline to produce patient-specific immunization recommendations automatically. This paper describes an overall approach, together with 3 specific strategies, developed to help deal with this problem. The paper then describes our experience applying the approach to a database containing over 400,000 immunization histories. The paper also discusses a number of the issues raised in adapting a computer-based guideline to accommodate incomplete patient data of this sort.

INTRODUCTION

A fundamental issue in the development of electronic medical record systems (EMRSs) concerns the quality of the data that are collected and stored [1]. Only if the data are accurate and in some sense complete, can they reliably be used to support patient care. This issue arises both 1) when clinicians interact directly with the EMRS, and 2) when computer-based clinical decision support tools make recommendations based on EMRS data.

Issues of data quality are also central in the current national initiative to build a broadly-based set of childhood immunization registries [2]. A particular data quality issue is that children's immunization histories are often incomplete in the sense that for one or more vaccine series it is known that certain previous doses were given, but the date of those doses may not be known. This frequently occurs because a child has moved from one geographic area to another or has switched from one provider to another within a given area.

The presence of incomplete vaccination histories (histories with missing dose dates) within an immunization registry database causes problems when attempting to provide patient-specific computer-based recommendations based on national immunization guidelines. The national guidelines are formulated assuming that the dates of all previous immunizations are known precisely. For example, they contain specific minimum wait-intervals between doses within each vaccine series and between doses of live vaccines of different series.

As discussed in detail later in this paper, there is no clear-cut approach to accommodating incomplete histories when formulating computer-based recommendations. This paper 1) describes three different strategies for handling incomplete histories that have been implemented in IMM/Serve, a computer-based guideline for childhood immunization, 2) discusses the circumstances in which each strategy might be appropriate, 3) describes preliminary experience with one strategy, and 4) discusses several issues that arise in confronting this problem. We believe that these issues are relevant not only to childhood immunization, but also potentially to other clinical domains, for example 1) where a sequence of interrelated actions needs to be taken over a period of time, or 2) where a guideline might usefully accommodate a certain degree of incompleteness or uncertainty in a patient's history.

IMM/SERVE: COMPUTER-BASED IMMUNIZATION FORECASTING

IMM/Serve is a computer-based forecasting program for childhood immunization [3]. It takes as input a child's immunization history and produces recommendations indicating which vaccinations are due and which should be scheduled next. IMM/Serve's recommendations are based on the nationally formulated recommendations of the CDC's Advisory Committee on Immunization Practices, but allows these recommendations to be customized to reflect local practice preferences.

IMM/Serve stores its domain knowledge 1) in tabular form (including the minimum age for each vaccine dose in different clinical circumstances and the minimum wait-interval between doses) and 2) in the form of if-then rules which encode the knowledge which determines which set of tabular parameters apply to a given case and which specific vaccine preparation should be used. A single version of IMM/Serve as a whole can contain several versions of each table and several variations of the rules [4]. This design allows IMM/Serve's logic to be customized on a case-by-case basis to accommodate the practice preferences of different clinics and health plans.

We have also built prototype tools to help maintain IMM/Serve's knowledge as the national recommendations evolve over time, and as the recommendations are customized to different local practice preferences [5, 6].

IMM/Serve is currently in operational use by the US Indian Health Service (IHS), initially in a clinic in Anchorage, Alaska with plans to extend the use to other IHS clinics nationwide during the coming year.

1091—8280/99/$5.00 © 1999 AMIA, Inc.

IMM/Serve is also being used on a test basis by the State of Oregon.

In bringing IMM/Serve into operational use, a number of important practical issues have surfaced, including issues involving data quality. A particular issue has been the prevalence of incomplete vaccination histories.

INCOMPLETE IMMUNIZATION HISTORIES

This section gives several examples of incomplete immunization histories to help make the various issues that arise more concrete. For simplicity, these examples involve only the DTP vaccine series. In practice, an incomplete history may have missing dose dates in several vaccine series.

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

In these example histories the dose number is in brackets following the date. The first three examples show different ways in which a "complete" history (a history with no missing dose dates) may appear in a registry database.

These three examples illustrate that different immunization registries may store different information for each dose. Some (see Hx-1) may have only the date of each dose. Some (see Hx-2) may always have the (presumed) dose number as well as the date. Some (see Hx-3) may have a mixture of doses with and without dose numbers. The latter situation may occur, for example, when a State registry combines immunization records from several different registry databases, some of which contain dose numbers and some of which do not.

Clearly, if no dose numbers are included then there is no way for a computer program to identify conclusively that a dose might be missing (although there might be patterns that suggested a possible missing dose). A missing dose can only be identified if at least one dose has a dose number of 2 or greater and is preceded by too few prior doses. Examples of such incomplete histories are shown below.

Hx-4: DTP - 1/1/97 [1], 5/1/97[3]
Hx-5: DTP - 5/1/97[3]
Hx-6: DTP - 3/1/97[2], 5/1/98[4]
Hx-7: DTP - 3/1/97, 5/1/97[3]
Hx-8: DTP - 6/1/97, 5/1/98[4]

These are examples of the type of incomplete histories that IMM/Serve is designed to accommodate. In Hx-4, dose 2 is missing. In Hx-5, doses 1 and 2 are missing. In Hx-6, doses 1 and 3 are missing. In Hx-7, one dose is missing (either dose 1 or dose 2). In Hx-8, two doses are missing (doses 1&2, 1&3, or 2&3).

The next two examples show one of the issues that arises when there is a mixture of doses with and without dose numbers.

Hx-9: DTP - 1/1/97 [1], 3/1/97[2], 5/1/97
Hx-10: DTP - 5/1/97[3], 7/1/98

The computer can only identify gaps prior to doses with dose numbers. As a result, if the computer is to process "mixed" histories, Hx-9 must be treated as a complete history with 5/1/97 counted as dose 3. Similarly in Hx-10, the first two doses are missing, and 7/1/98 must be treated as dose 4. As described later in the paper, two of IMM/Serve's strategies allow processing of mixed histories of this sort.

Hx-11: DTP - 1/1/97 [1], 3/1/97[2], ?/?/?[3]

A final possible type of incomplete history, occurs when the clinician determines from talking to the child's parent that the most recent vaccination has been given, but does not know the precise date. Here a computer cannot be used to apply the guidelines, and the clinician must use clinical judgement to determine when the next dose is due.

STRATEGIES FOR HANDLING INCOMPLETE IMMUNIZATION HISTORIES

In adapting IMM/Serve to accommodate incomplete histories, our overall philosophy was to make as few changes as possible to the existing knowledge base (KB), the tabular and rule-based guideline knowledge. As much as possible, we wanted only to modify the underlying inference engine that operated upon the KB. We discuss our rationale for this decision later in the paper.

This section first describes IMM/Serve's overall approach to handling incomplete histories. It then describes different specific strategies that can be used to implement this approach.

Overall Approach

To accommodate missing dose dates without changing the KB logic, the following logic was incorporated into the C program that coordinates IMM/Serve's analysis. In a sense, the approach involves "manufacturing" computer-generated doses that "fool" the KB logic into operating correctly. After describing each step, we discuss its rationale.

1. If a history has missing doses, then manufactured doses are added at the first available date. For example, if dose 1 is missing, a dose is added immediately after the birthdate. If doses 2 and 3 are missing, two doses are added on the two successive dates after dose 1. These doses are flagged as "manufactured."

   Rationale: The manufactured doses are inserted to provide a complete history so that the rest of IMM/Serve can operate. It is easier to choose arbitrary dates than to try to determine valid dates for the missing doses. As described below, IMM/Serve adjusts its processing of these manufactured dates appropriately.

2. The new history (containing manufactured doses) is then input to IMM/Serve. In its analysis, IMM/Serve has been modified to give the manufactured doses special treatment (by disabling certain tests). For example, if two doses are too close together, the second dose normally is flagged as "invalid" and not counted as part of the history. The required intervals are stored in a "dose screening table." If the second dose is manufactured, however, this test automatically returns "ok", the manufactured dose is accepted as valid, and processing continues.

   Rationale: Since we know that we have given inappropriate dates to the manufactured doses, we need to disable the checking of minimum ages and wait-intervals for those doses.

3. The dose screening table previously contained minimum wait-intervals only between successive doses. To handle gaps in the history, minimum wait-intervals were added to the dose screening table specifying for each dose of each vaccine series how long to wait from each and every previous doses in that series. (This was the only information added to the KB to accommodate missing doses.) As a result, if DTP dose 4 is being evaluated and both DTP dose 2 and DTP dose 3 are manufactured, the screening logic would look back to DTP dose 1 and make sure that enough time had passed between doses 1 and 4. If not, dose 4 would be considered invalid.

   Rationale: This assures that there is a sufficient interval for valid missing doses between the real doses.

4. If at any time in the process of analyzing a vaccine series history containing manufactured doses, a real dose is screened as being invalid, then analysis of that vaccine series halts with an error message. Analysis of the other vaccine series for the patient, however, is still performed.

   Rationale: As described later in the Discussion, it is important to place limits on the degree of incompleteness and uncertainty that IMM/Serve will attempt to accommodate.

Due to the nature of the immunization logic, this approach allows missing doses to be handled with minimal modification to the KB (the tabular and rule-based knowledge). The remainder of this section describes specific strategies that IMM/Serve uses to implement this overall approach.

Default Strategy (Ignoring Dose Numbers)
Before describing IMM/Serve's three strategies for handling incomplete histories, it is worth pointing out that a simple default strategy is simply to ignore dose numbers and process an immunization history sequentially by date. This is how IMM/Serve was originally designed to be used, and is the only viable option if no dose numbers exist. The strategy may also be useful, for example, when summary statistics for a population of patients are desired and it is believed that few doses are missing and/or that dose numbers may be unreliable.

Mandatory Dose Number Strategy
This strategy was implemented to meet the needs of the IHS, which stores a dose number for each dose, making the presence of missing doses easy to identify. Here IMM/Serve first checks that each dose of a vaccine series has a valid dose number and that the corresponding dates are in sequence and otherwise make sense. It any doses are missing, manufactured doses are created and analysis proceeds as described above.

Low Dose Number Strategy
This strategy was developed to meet the current needs of the Oregon immunization registry which in 1998 had over 400,000 patient records created by combining records from several sources, and which contained a significant level of data error. In a given vaccine series, some doses have dose numbers and some do not. In this strategy, the lowest numbered dose (if any) is treated as correct. If any previous doses are missing, manufactured doses are created. Any subsequent doses are processed ignoring dose numbers. This approach has allowed Oregon to run IMM/Serve on its records to help test IMM/Serve itself, and to help identify potential errors in the data.

Designated Dose Strategy
In the designated dose strategy, a user must indicate for each vaccine series with missing doses, a particular dose which should be considered correctly numbered. Analysis then proceeds in a fashion analogous to the low numbered dose processing described above. The designated dose approach is not currently being used operationally, but was developed to allow a user to exert some control over the processing of an incomplete history.

PRELIMINARY ASSESSMENT OF THE APPROACH

This section describes a preliminary assessment of our approach using a 1998 version of the immunization registry database of State of Oregon which contains data on 431,024 children. This database was constructed by merging data from several different sources, including private providers, health maintenance organizations, insurance companies, and county immunization registries. The database is currently being refined to remove duplicate records, as well as to remove duplicate vaccinations within records. The database contains a mixture of doses with and without dose numbers, and was processed using IMM/Serve's low dose number strategy described above.

In a separate project, we have developed a pilot computer-based tool (IMM/Scrub) to assist in the deduplication of the immunization history records. To perform our assessment, we first ran IMM/Scrub to remove as many redundant doses as possible. We then analyzed the resulting cleaned immunization history data for each vaccine series for each patient to determine how many incomplete series histories were present. As described previously, a vaccine series history was considered incomplete if a numbered dose had too few prior doses (e.g., if Hib dose number 3 had only one prior Hib dose listed).

Of the 431,024 cases, 60,336 (14%) had one or more incomplete series history. These were broken down by series as follows:

Series Name	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 1) that the presence of incomplete histories is indeed a significant problem, 2) that cases frequently have incomplete histories in more than one series (since 60,336 cases contain 97,744 incomplete series histories), and 3) that an approach to computer-based immunization forecasting that successfully accommodates such histories could be very helpful. The following table indicates, for each series with an incomplete history, how many (and what percentage) of those series could be successfully run using IMM/serve's low dose number strategy.

Series Name	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%)

All but 17,307 (18%) of the vaccine series with incomplete histories were able to be run by IMM/Serve. Most of the failures (16,822) were due to data quality problems: 1) 6,491 series had a dose number that was too high for the series, 2) 10,134 series had dose numbers which were not in chronological order, 3) 8 series had too many prior doses before a numbered dose (despite meeting the incomplete history definition described above), and 4) 189 series had some other data quality problem.

For 430 series, IMM/Serve was unable to insert manufactured doses into the incomplete history because there was an insufficient time interval. For 55 series, IMM/Serve had to screen a real dose as invalid after inserting its manufactured dose(s) into the series history. As described previously, in this circumstance IMM/Serve aborts its analysis of the series.

These figures indicate that IMM/Serve was able to handle all the incomplete series histories in an appropriate fashion. In only 55 series (from 431,024 cases) was the situation so undefined that IMM/Serve had to abort its analysis. (When it does this, it produces an error message, and leaves the clinician to decide how to handle that series.)

DISCUSSION

This section discusses some of the interesting issues that arose in the process of formulating IMM/Serve's approach for handling incomplete immunization histories.

The Need for Different Strategies

A number of factors may influence the most appropriate strategy for dealing with incomplete immunization histories for a given registry. These factors include the policies, structure and maturity of the registry and of its database. If a registry does not store dose numbers, then the issue is not relevant. If a registry is mature in the sense that it is well organized, has a database containing accurate data which always includes dose numbers then the "mandatory dose number" strategy is appropriate. If some doses have dose numbers and some do not, then 1) the "designated dose" strategy gives the user a degree of control in the process, and 2) the "low dose number" strategy may facilitate initial analysis of the database records in the presence of significant degree of inaccurate data. As we obtain additional practical experience working with different registries, we may find that additional variations on these strategies will be more appropriate in certain circumstances.

Handling Incomplete Histories by Augmenting the KB vs. Modifying the Inference Engine

As described above, we have taken the approach of making as little modification as possible to IMM/Serve's KB (the tabular and rule-based knowledge), and making almost all the modifications to the inference engine which operates on the KB (a C program). An alternative approach would have been to extend the logic in IMM/Serve's KB to handle all the different combinations of clinical conditions that might arise with incomplete histories. We felt that this approach would have been difficult for several reasons. Augmenting the KB to deal with all the new combinations of clinical conditions could have vastly expanded that logic, especially if several strategies were to be used. Such augmentation would also have made the KB much more difficult to maintain, both because of its additional complexity and because it would be very difficult to get domain experts to help define all the complex interactions involved.

The approach we have taken leaves the KB unchanged, so that it expresses the logic as required when all dates are known. It is fortunate that the nature of the immunization domain is such that this simplification is possible. In other domains where this type of problem arises, there may not be an easy way to modify the inference engine while leaving the KB essentially unchanged.

Placing Limits on the Allowed Degree of Incompleteness

An important element in implementing the approach was developing a set of well-defined conditions in which the approach would be applied. One example of this is seen in the "mandatory dose number" strategy, where IMM/Serve requires that all doses numbers be specified. If this does not occur, analysis of the vaccine series is terminated with an error message. A second example is seen if manufactured doses are inserted into a vaccine series history, but the screening logic later indicates that a real dose is invalid. Here again, IMM/Serve terminates its analysis with an error message rather than attempt to deal with all the potential complexity that might arise trying to determine how best to handle this situation.

Data Cleaning vs. Immunization Forecasting: Where to Draw the Line?

Data cleaning is the process of making the data in a database accurate, non-redundant, and complete. The goal is not necessarily to make all data elements be perfect, but rather that the database be "fit" for its intended use. In operating an immunization registry, the process of data cleaning is normally considered to be separate from the process of running a forecasting program to produce patient-specific recommendations. As described in this paper, however, these two processes need not be completely distinct. An immunization forecasting program can be designed to accommodate certain types of data incompleteness and uncertainty.

SUMMARY

This paper has described how a computer-based guideline which produces patient-specific recommendations can accommodate certain types of missing data in a patient history. Over time we anticipate that additional variations of the strategies described will prove useful to other registries. It is also interesting to speculate that computer-based guidelines in other clinical domains might also gain enhanced utility by being modified to accommodate certain types of incompleteness and uncertainty in a patient's history. The current project is one step in exploring how this might be accomplished.

Acknowledgments

This work was supported in part by cooperative grant U1W/CCU114707 from the Centers for Disease Control and by NIH grant R01 LM06682 from the National Library of Medicine.

References

1. Hogan WR, Wagner MW: Accuracy of data in computer-based patient records. JAMIA 1997; 4:342-355.
2. Cordero JF, Guerra FA, Saarlas KN. (Eds.): Developing immunization registries: Experience from the All Kids Count Program. Am J Preventive Med 13 (Supplement): entire issue, 1997.
3. Miller PL, Frawley SJ, Sayward FG, Yasnoff WA, Duncan L, Fleming D. Combining tabular, rule-based and procedural knowledge in computer-based guidelines for childhood immunization. Computers and Biomedical Research 1997; 30:211-231.
4. Miller PL, Frawley SJ, Sayward FG: Issues in accommodating national changes and local variation in a computer-based guideline for childhood immunization and in related knowledge maintenance tools. Proceedings of the 1998 AMIA Annual Fall Symposium, Orlando FL, November 1998, pp. 563-567.
5. Miller PL. Tools for immunization guideline knowledge maintenance I: Automated generation of the logic "kernel" for immunization forecasting. Computers and Biomedical Research 31:172-189, 1998.
6. Miller PL, Frawley SJ, Brandt C, Sayward FG. Tools for immunization guideline knowledge maintenance II: Automated Web-based generation of user-customized test cases. Computers and Biomedical Research 31:190-208, 1998.