Introducing an Academic Data Warehouse into the Undergraduate Medical Curriculum

Jason A. Lyman, MD, MS, Wendy Cohn, PhD, William Knaus, MD, Jonathan S. Einbinder, MD, MPH

Department of Health Evaluation Sciences, School of Medicine
University of Virginia, Charlottesville, Virginia

ABSTRACT

There is increasing interest in integrating population health and informatics topics into the undergraduate medical curriculum, yet little consensus exists on the most effective approach to accomplish this. We introduced the use of an academic data warehouse of encrypted patient information into an existing 2nd year medical school course. Exercises were developed requiring students to retrieve and interpret information regarding local disease prevalence, practice patterns, and patient characteristics. These exercises were integrated into existing weekly problem sets in a multiple-choice format. Faculty and student perceptions were assessed with surveys, and augmented with interviews of student volunteers, and database usage statistics. Our results indicate widespread agreement among both students and faculty that population-based medicine warrants inclusion in undergraduate medical education. The majority of the students felt the exercises complemented the clinical cases around which they were structured. There was less agreement, however, that the exercises were valuable, with several students suggesting a more open-ended, discussion-oriented approach. It was clear that faculty perceptions had a significant impact on student reactions.

INTRODUCTION

In recent years, there has been increasing recognition that population health and informatics topics are under-represented in undergraduate medical curricula, though there is evidence that this is changing.

In 1998, the Association of American Medical Colleges (AAMC) published its initial report outlining consensus-based learning objectives for undergraduate medical education, as part of the Medical School Objectives Project (MSOP). Medical students, it states, should demonstrate “knowledge of the epidemiology of common maladies within a defined population...”, with skills that include “the ability to retrieve (from electronic databases and other resources), manage, and utilize biomedical information for solving problems and making decisions”. In Report II of the MSOP, population health and medical informatics were specifically addressed, with an emphasis on improving skills in information retrieval and management.

At the University of Virginia Health System (UVaHS), we modified a second-year medical school course to include information management and population health topics, using our Clinical Data Repository (CDR), a patient data warehouse at the UVaHS. Here we describe the course modifications, and a subsequent evaluation was performed involving both students and faculty members.

METHODS

The Practice Of Medicine 2 Course

Practice of Medicine 2 (POM 2) is a second-year medical student course at the UVa School of Medicine that focuses on clinical problem-solving, diagnostic approach, pathophysiology, and treatment. It augments didactic lectures with small group sessions, during which students discuss clinical cases.

The course is divided into fourteen sections (e.g. Gastroenterology, Infectious Disease), each lasting one to three weeks. Each section has a section leader who generates weekly problem sets for the students. These problem sets are composed of a series of cases followed by multiple-choice questions targeting diagnosis and treatment issues. These are turned in at a two-hour specialist session that occurs every Friday. These sessions are composed of 15-20 students and run by a clinical faculty member with specialty training in the section being covered.

The course also includes a two-hour “tutorial” group session, composed of six students and one clinical faculty member. This “tutor” is assigned to a specific group of students for the entire one-year course. During these sessions, one or more cases are discussed, and learning objectives are developed by each student to be researched for the following week.
**Incorporating Use Of The CDR**

The CDR is a relational data warehouse that integrates UVaHS patient data from multiple legacy systems, including administrative, billing, and laboratory information systems. Initially developed primarily for research, it has since evolved as a tool for medical management, as well as medical student and housestaff education.

In 2000, the UVa School of Medicine Curriculum expressed its commitment to incorporate population health and information management topics into the medical school curriculum, with the recognition that the CDR had significant potential as an educational resource. In the summer of 2001, two informaticists (JAL, JE) in the Department of Health Evaluation Sciences met with the POM 2 course directors to discuss options for integrating the use of the CDR into the course. The decision was made to include CDR exercises in the weekly problem sets, which would be evaluated by the specialist small group leaders. These faculty members would be provided with answer sheets to facilitate discussion.

An informatician on the CDR project team developed the CDR exercises [JAL]. An effort was made to develop questions that provided clinically relevant data that complemented the cases in the problem set. When feasible, questions were suggested and/or reviewed by the section leader. Typically, the problem set cases would be reviewed for areas in which a population-based query might provide interesting information.

The students who began POM 2 in the academic year 2001-2002 had received previous instruction and experience with the CDR in a first-year medical student course. A lecture early in the academic year was planned to review the use of the CDR, as well the goals for its use in the POM 2 course. The CDR was emphasized as a resource where students could obtain data about local disease prevalence, practice patterns, and patient characteristics.

For support, CDR “office hours” were established, during which a CDR project team member would be available in the library for two hours per week.

**Evaluation**

In February of 2002, we surveyed faculty and students to elicit their perceptions of the integration process and identify areas for improvement. Student and faculty surveys were developed, and reviewed by an expert in program evaluation (WC).

The anonymous paper-based student survey was handed out in class and subsequently e-mailed as an attachment to the entire class. The survey included both Likert scale multiple-choice questions as well as open-ended questions. The faculty survey was Web-
based, and both section leaders and specialist group leaders who had reviewed CDR exercises with the students were e-mailed a link to the survey web page. These surveys were also anonymous.

In addition, student volunteers were identified to participate in brief interviews. These interviews were conducted by one of the authors [WC] with no role in the POM 2 course or the CDR project team, and the identity of the interviewees was kept confidential from all other investigators.

RESULTS

CDR Exercises
Fifteen exercises were assigned between 9/1/2001 and 2/22/2002, for the Pulmonary, Allergy, Cardiology, Infectious Disease, Rheumatology, Obstetrics-Gynecology and Renal sections. Students were asked to explore the CDR to learn about such topics as nephrotic syndrome, tuberculosis, asthma, and acute myocardial infarction. The exercises evolved over the course of the semester from simple queries asking about specific numbers to more open-ended questions requiring the students to make inferences from the data (Figure 1). For each exercise, the specialist small group leaders received a 1-2 page answer sheet, containing relevant data and points for discussion.

Student Usage
We reviewed CDR usage statistics for the 143 POM 2 students from 9/1/2001 to 2/23/2002, which revealed a median number of logins of 21, with a mean, standard deviation, and range of 18.8, 11.6, and 0-72 respectively. Of note, twenty-two students had no logins recorded during that time period.

Survey Responses
Survey response rates are shown in Table 1. The Likert scale survey questions and response distributions for students and faculty are shown in Tables 2 and 3 respectively.

In addition, students felt confident in their ability to submit queries and answer questions: 84% indicated they were Very Confident or Extremely Confident in their ability to formulate queries, and 64% responding similarly regarding their ability to find relevant data in the report menu. When asked about what aspect of using the CDR was easy, students responded with comments such as “The whole CDR is pretty easy to use” and “pretty simple”. Asked to comment on aspects that were difficult, many students referred to the challenge of identifying appropriate diagnosis codes for query formulation: “finding which Dx is appropriate and deciding whether to use CCHPR or ICD9 codes”.

Survey responses indicated that the exercises were not particularly time consuming, and the mean, standard deviation, and range for time spent per question was 11.6, 7.8, and 2-45 minutes respectively. Several suggested that database performance was an issue with one user expressing frustration “when server is busy or query is long”.

When asked to suggest improvements to the CDR integrations, students had several ideas: “I think the

<table>
<thead>
<tr>
<th>QUESTION TEXT</th>
<th>Strongly Agree</th>
<th>Somewhat Agree</th>
<th>Neutral</th>
<th>Somewhat Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population-based medicine should be incorporated into the medical school curriculum.</td>
<td>22 (49%)</td>
<td>22 (49%)</td>
<td>1 (2%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>The goals for incorporating use of the CDR into the POM 2 course were clearly presented.</td>
<td>8 (18%)</td>
<td>14 (31%)</td>
<td>14 (31%)</td>
<td>6 (13%)</td>
<td>3 (7%)</td>
</tr>
<tr>
<td>In general, the CDR questions complement the cases in the problem sets, conveying relevant and useful information.</td>
<td>11 (24%)</td>
<td>18 (40%)</td>
<td>6 (13%)</td>
<td>6 (13%)</td>
<td>4 (9%)</td>
</tr>
<tr>
<td>The CDR exercises take an appropriate amount of time to complete.</td>
<td>11 (24%)</td>
<td>16 (36%)</td>
<td>11 (24%)</td>
<td>2 (4%)</td>
<td>5 (11%)</td>
</tr>
<tr>
<td>The CDR questions are a valuable component of the weekly problem sets.</td>
<td>3 (7%)</td>
<td>9 (20%)</td>
<td>15 (33%)</td>
<td>11 (24%)</td>
<td>7 (16%)</td>
</tr>
<tr>
<td>The level of support available for the CDR has been adequate.</td>
<td>20 (44%)</td>
<td>11 (24%)</td>
<td>11 (24%)</td>
<td>2 (4%)</td>
<td>1 (2%)</td>
</tr>
</tbody>
</table>

Table 2: Student Survey Responses
CDR is a valuable tool and that it is important for us to know how to use it. However ... I think every other week would be more appropriate for CDR questions”, and “maybe formulate more open-ended questions that can be used as a discussion topic rather than a multiple choice”.

The faculty respondents who supplied comments expressed lack of familiarity with the CDR, “I have never used CDR...I found myself teaching the medicine, but unable to discuss the process.”; interest in gaining access, “Any way to expedite our getting onto this?”; as well as suggestions for improvement, “decrease student busy work. Teach them how to search etc then provide example searches already done by you .... This will teach them the value but limit the work in an already loaded curriculum”.

Student Interviews
Five students participated in confidential interviews lasting about forty-five minutes. In general, the interviews revealed similar issues to the survey results. These students were positive about the CDR, though the tasks were sometimes perceived as mundane: “I like doing the CDR exercises...it hits home when you see the numbers in terms of population based medicine. It seems like busy work but it is important to learn the tool.” Specific suggestions were made to improve the CDR integration: “Change Friday problems to case discussion; participation is your grade...” One important theme that emerged was the variety of faculty attitudes toward the CDR, and the impact this had on student perceptions: “[mentors provide] little to no discussion of what the CDR exercises mean”, and “One mentor does assign learning objectives that have to do with the CDR. Her desire to use it motivated us”

### DISCUSSION

Several important issues arose during both development and evaluation. Technical issues affected question development and student satisfaction. The evaluation highlighted the importance of population health, the format of the CDR exercises, and the need for faculty support.

**Technical Issues**

As CDR exercises were developed, it often became apparent that changes in the user interface of the database were necessary to optimize access to useful information. Students, asked to explore patient demographics, disease epidemiology, and treatment patterns, need summarized data. The creation of reports for admission laboratory results and demographic summaries, for example, were motivated by the needs of these new users. In addition, the addition of over 140 new regular users presented challenges related to database performance. Prior to this effort, most of the data loading was done overnight, times when students often performed their queries. Some students indicated frustration with performance issues related to these factors.

**Evaluation**

Important issues that arose from the evaluation included the rationale for integration, format of the exercises, faculty attitudes.

**Rationale**

Despite the fact that the overwhelming majority of respondents felt strongly that population health deserved inclusion in undergraduate medical education, there was evidence that the rationale for the use of the CDR in the POM 2 course was unclear. Only one lecture was provided to the students that addressed this issue, and there was little direct contact with the specialist group leaders, both of

<table>
<thead>
<tr>
<th>QUESTION TEXT</th>
<th>Strongly Agree</th>
<th>Somewhat Agree</th>
<th>Neutral</th>
<th>Somewhat Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The goals for having the CDR questions integrated into the problem sets are clear.</td>
<td>4 (22%)</td>
<td>5 (28%)</td>
<td>6 (33%)</td>
<td>2 (11%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Population-based medicine topics should be incorporated into the medical school curriculum.</td>
<td>10 (56%)</td>
<td>6 (33%)</td>
<td>2 (11%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>In general, the CDR questions complemented the cases.</td>
<td>4 (22%)</td>
<td>6 (33%)</td>
<td>3 (17%)</td>
<td>3 (17%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>The CDR questions focused on interesting and relevant aspects of the diseases being studied.</td>
<td>1 (6%)</td>
<td>5 (28%)</td>
<td>4 (22%)</td>
<td>3 (17%)</td>
<td>1 (6%)</td>
</tr>
<tr>
<td>The CDR questions were a valuable component of the weekly problem sets</td>
<td>3 (17%)</td>
<td>3 (17%)</td>
<td>5 (28%)</td>
<td>4 (22%)</td>
<td>2 (11%)</td>
</tr>
<tr>
<td>The CDR answer sheets were helpful.</td>
<td>4 (22%)</td>
<td>6 (33%)</td>
<td>3 (17%)</td>
<td>3 (17%)</td>
<td>1 (6%)</td>
</tr>
</tbody>
</table>

Table 3: Faculty Survey Responses

477
which may have led to this lack of clarity.

Format of the CDR Exercises
Creating CDR exercises in a multiple-choice format facilitated integration into the problem sets, but respondents expressed a frustration with this approach. Moving the CDR questions to the weekly tutorial sessions and making them more discussion-oriented were frequent recommendations. A limitation of this approach is the reliance on the mentors to be able to lead such a discussion.

Faculty Attitudes and Support
Our approach depended upon the specialist small group leaders to review their CDR answer sheets in advance of the session, evaluate the students’ responses, and discuss the implications of their findings. It was clear that this did not often occur. Equally evident was the significant impact that faculty attitude can have over the students’ reactions: one clinical tutor who had students use the CDR for weekly learning objectives helped the students understand the importance of such information.

Limitations
The generalizability of our findings to the entire 2nd year class of students and all POM 2 faculty are limited by our response rates. The information derived from our evaluation, however, is helpful in identifying specific areas for improvement and potential barriers that need to be addressed.

Future Directions
We will use the result of this evaluation to guide our next steps in incorporating population health topics into the curriculum. Specifically, we will consider reducing the number of CDR questions from the problem sets and trying to incorporate their use in the more flexible tutorial sessions. One of the benefits of this approach is that there are fewer faculty tutors, which may make it easier to provide them with information regarding the CDR and give them a more participative role in the integration.

We will also explore the option of using other courses to incorporate these topics, such as Clinical Epidemiology, a 2nd-year course that might provide a natural setting for discussion of these topics.

CONCLUSION
We utilized an academic data warehouse of encrypted patient data as a vehicle for the integration of population health topics in a 2nd year medical student course. We were encouraged that the majority of the student respondents felt that using the database provided interesting and relevant information that complemented the cases, but clearly more effort is needed in identifying the optimal approach for “fitting” the CDR into the curriculum. It is similarly clear that the level of faculty support for such an initiative is critical, and can have a large impact on students’ reaction regarding the utility and relevance of such an information resource.

ACKNOWLEDGMENTS
The authors gratefully acknowledge the assistance of the assistant course director for POM 2, Ms. Darci Lieb.

REFERENCES