A taste of individualized medicine: physicians’ reactions to automated genetic interpretations

Hallvard Lærum, Sara Bremer, Stein Bergan, Thomas Grünfeld

ABSTRACT

The potential of pharmacogenomics is well documented, and functionality exploiting this knowledge is about to be introduced into electronic medical records. To explore physicians’ reactions to automatic interpretations of genetic tests, we built a prototype with a simple interpretative algorithm. The algorithm was adapted to the needs of physicians handling immunosuppressive treatment during organ transplantation. Nine physicians were observed expressing their thoughts while using the prototype for two patient scenarios. The computer screen and audio were recorded, and the qualitative results triangulated with responses to a survey instrument. The physicians’ reactions to the prototype were very positive; they clearly trusted the results and the theory behind them. The explanation of the algorithm was prominently placed in the user interface for transparency, although this design led to considerable confusion. Background information and references should be available, but considerably less prominent than the result and recommendation.

OBJECTIVE

The objective of this study was to develop a prototype for automated interpretation of genetic tests, and evaluate hospital physicians’ reactions to it in a specific use case. The use case provided guidance for immunosuppressive treatment during kidney transplantation based on the CYP3A5 gene.

BACKGROUND AND SIGNIFICANCE

The use of genomics in clinical practice constitutes an integral part of individualized medicine, and is a common goal for numerous practitioners and researchers.1 2 Despite its potential, the translation of pharmacogenomics into clinical practice has been slow.3 There seems to be a gap in physician’s knowledge about pharmacogenomics4 and a low rate of genetic tests.5 Also, the rate of growth in genetic knowledge seems to outperform physicians’ ability to acquire it.6 Solutions for automated interpretation of genetic variations are already being tested, but there are few studies investigating how physicians react to them.6–8 Earlier attempts at providing physicians with expert advice showed mixed results,9 partly because of lack of transparency of the algorithms producing the advice.10 In this study, we wanted to investigate how physicians would react to ready-made genetic interpretations made by a system. These interpretations often stem from knowledge that may be new to the physicians and first encountered in this system. We wanted to study how such information should be presented to achieve the necessary trust of the physicians and avoid being perceived as a threat to their professionalism.11

MATERIALS AND METHODS

Development of the prototype

Prior to development, the work process of physicians relevant to the use case was described and analyzed in a standardized, structured way.12 The research group agreed on a written requirement description and generic design. The database model was based on HL7 Domain Information Model for Genetic Investigations.13 The application allowed the user to select an algorithm that produced a report interpreting the genetic variants of the patient for a given subject (see online appendix 1 for a description of the application).

The pharmacologists in the team developed an algorithm that interpreted variants of the CYP3A5 gene and their impact on the metabolism of the immunosuppressive drug tacrolimus. This drug is used to avoid organ rejection after renal transplantation. Patients with specific CYP3A5 gene variants may require twofold higher tacrolimus doses than patients without these gene variants.14–17 The CYP3A5 case was selected because it is well documented, not yet used by clinicians, relevant to their work, and the test was available at our hospital. The pharmacologists also wrote the description of the results, recommendation, detailed interpretation, and logic of the algorithm in the application.

The investigation

The impact of the application was investigated by an adapted ‘think aloud’ technique,18 in which the user was asked to vocalize thoughts and considerations during use of the application. The physicians’ speech and actions on screen were recorded, and they were observed while resolving two patient scenarios with regard to tacrolimus treatment. In the scenarios, the physicians were asked to identify the correct dosing for a given patient, based on information provided by the prototype. They responded to a web questionnaire about general attitudes to genetic testing adapted from Bonter et al5 before the scenarios and questions about reactions to the application after the scenarios. Finally, a short free-form interview was recorded to obtain responses and thoughts not covered by the questionnaire. A list of the software used in the development and investigation can be found in online appendix 2.

Nine physicians were recruited for the investigation, six of 12 possible from the Section Of Transplantation Surgery, and three of 16 possible from the Section Of Nephrology, both sections at the Department Of Transplantation Medicine at Rikshospitalet, Oslo University Hospital. The surgeons were selected randomly, and the internists were selected by convenience sampling performed.
by the head of the section. The sample consisted of five junior doctors and four senior doctors, and the male/female ratio was 2:1. All the physicians were using tacrolimus in their daily practice, but the CYP3A5 test was not part of their kidney transplantation protocol.

RESULTS
Demographics and competence
The group of respondents represented both experienced and less experienced physicians. Three of the nine physicians had been working for less than 12 months at the department. Five of the physicians had completed a specialty after qualifying as a medical doctor, two of them more than 6 years previously. One physician had a PhD in molecular genetics, but the others had no special training in medical genetics. Most of the physicians had little experience with ordering genetic tests, and only two had ordered pharmacogenetic tests or the CYP3A5 test before (see online appendix 3). On the basis of subjective observations, eight of the nine physicians appeared to be efficient users: most navigations took less than a second, and none of the participants needed instruction on the basic aspects of the user interface.

Attitudes to genetic testing
In general, the physicians were positive about genetic testing and incorporating it into their clinical practice. They were mostly positive about the use and impact of the genetic test in their own practice, but not all were comfortable discussing such tests with the patients. As a group, they were not confident in their ability to interpret genetic tests themselves (table 1).

Use of the prototype
The physicians were observed while testing the prototype for two patient scenarios. The median time to resolve scenario 1 was 164 s (range 110–339). In this scenario, seven of the nine physicians used more than 15% of their time studying the description of the logic of the algorithm (median 25.3%, range 7.2–51.5%). Eight of the nine physicians were observed not to immediately grasp the concept of the ‘interpreted report’ versus the original genetic data. This led to widespread hesitation, and many requested help before moving on. Four of the nine physicians were observed to confuse the description of the logic of the algorithm with the result of the algorithm, and this led to one erroneous response for scenario 1. The second scenario was resolved in less time with no hesitation (median duration was 41.8% of that for scenario 1), and there were no errors.

Attitudes to the application
Despite the limited scope of the system, the physicians were very positive about it. In the survey, they were positive about both the system in general and how it worked. During the observations, five physicians spontaneously made positive remarks about the system (citations 1–5 in online appendix 4; table 2).

The physicians also commented on the various parts of the system: four found the details section too extensive, unclear, or difficult to understand (see online appendix 3). In the transcriptions of the video recordings, all physicians expressed confidence in the content, and showed varying interest in investigating the references (citations 8 and 9 in online appendix 4). When discussing the layout, one physician expressed a clear preference for recommendations (citation 14). Several other physicians expressed similar opinions (citations 13, 15 and 16). None of the physicians checked the references before answering the scenario questions.

When the physicians were asked about further development of the system, they said that such a system need not cover all conceivable genetic use cases to be useful. They were positive about the automatic interpretations, regardless of the authors being external or the genetic test being mentioned or not in the clinical protocols. They were unclear whether or not they had the time to study genetics in addition to their own specialty, but they felt that the reports made them more interested in the subject (see online appendix 3). The physicians were also asked to prioritize further expansions of the prototype. Of 11 suggestions, the highest prioritized expansions were integration into the electronic medical record and more report types.

DISCUSSION
This is a small study, using multiple data sources and methods to demonstrate physician reactions to a relevant but limited prototype. The results do not describe a physician population, but provide a valuable presentation of a typical use case. The selection of participants was deemed representative of those initiating tacrolimus treatment (surgeons), but not necessarily of physicians continuing the treatment (internists). The test situation contained only simulated data and decisions, making the

<table>
<thead>
<tr>
<th>Table 1 Attitudes to genetic testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree completely</td>
</tr>
<tr>
<td>Can influence treatment decisions</td>
</tr>
<tr>
<td>Can improve outcomes</td>
</tr>
<tr>
<td>Sufficient evidence to order</td>
</tr>
<tr>
<td>Feel sufficiently informed</td>
</tr>
<tr>
<td>Able to interpret results</td>
</tr>
<tr>
<td>Comfortable discussing genetic test results with patients</td>
</tr>
<tr>
<td>Patients are asking about genetic tests</td>
</tr>
<tr>
<td>Patients fear discrimination based on genetic test results</td>
</tr>
<tr>
<td>Has influenced treatment plans</td>
</tr>
<tr>
<td>Has increased diagnostic benefit</td>
</tr>
<tr>
<td>Has increased therapeutic benefit</td>
</tr>
<tr>
<td>Has identified predisposition/high risk</td>
</tr>
</tbody>
</table>
situation less realistic. The presence of the observer may have increased the pressure, making the situation similar to a clinical setting in some respect. Subjectively, the majority of the physicians seemed to be efficient users, and this may have led to a more positive attitude towards the system. However, physicians in Norwegian hospitals have been using electronic medical records for almost a decade, making the expected proportion of inexperienced users low.19

If the physicians already had and trusted the knowledge presented in the system, this could have influenced the responses about the system itself. This could also have reduced the need to check the background of the report, explaining some of the observed behavior. Indeed, most of the physicians were aware that tacrolimus metabolism had a genetic component. Also, they generally had greater confidence in genetic tests than found in other studies.5 20 However, the genetic test was not incorporated into their daily practice, and most of the physicians responded that they were not able to interpret genetic results themselves. This suggests that the system represented at least some new knowledge.

The study does not rule out the possibility that the positive attitudes could have partly been due to selection bias of the three internists, in that those interested in computers might have been selected for the study. The attitudes may also have been due to selection bias of the knowledge selected for the algorithm in the prototype, as it shows a very good match to the physicians’ needs and is very well documented.14 However, in pharmacogenomics, the evidence supporting similar advice for

<table>
<thead>
<tr>
<th>Table 2 Evaluation of the prototype system</th>
<th>Disagree completely</th>
<th>Disagree partly</th>
<th>Disagree partly</th>
<th>Agree completely</th>
</tr>
</thead>
<tbody>
<tr>
<td>The information I get from the system is useful</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>I don’t understand the purpose of this system</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>The system is promising</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>The interpretation in the reports is trustworthy</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>The knowledge in the system appears outdated</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>This is useful for my daily routine</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>The usefulness of the system makes it worthwhile logging into</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>The system is of no use until it is integrated into the electronic medical record</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>The system is hard to use</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I easily found out how to use the system</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>I need training in this system</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>The system is sufficiently fast</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>The information was unclearly presented</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>The content was nicely organized; it was easy to find what I needed</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 1 The ‘Patient summary’ and ‘Create report’ windows of the application.
other medications is considerable, and the CYP3A5 gene is one of many feasible candidates for automatic interpretation in the future.

Response bias due to the presence of the investigator is also possible. The respondents knew that the prototype and its knowledge were produced by local colleagues. Their trust in us may have been extended to the system, or led to self-restriction of their answers. However, the responses were both positive and negative, and the observed comments and behavior were consistent with the answers in the web survey.

The observations revealed that several physicians were confused by the description of the algorithm used to produce the reports, and this led to misinterpretation on one occasion. To increase transparency, an algorithm description was displayed in the original design (figure 1), but it appeared that the physicians expected a result instead. The physicians in our study expressed a preference for seeing the results and recommendation first, then the explanations, and finally the references, if needed. This fits well with the structure of the genetic HL7 CDA format, but also with the usability concerns recently proposed by Bushell.

Both the explanations and references in genetic reports are needed for practical reasons and to build confidence, but they should not overshadow the results needed in daily practice. The physicians stated that the references provided an easily accessible option for those who look for it.

The prototype appears to have offered the physicians information that they found useful. Such information should be offered as decision support in the electronic medical record—for instance, in provider order entry or in treatment protocols. We believe that clear, preformed algorithms achieve trust more easily than complicated ‘black box’ type decision support, as the former is easier to explain and document. Early versions of such functionality probably need not cover all drugs with possible genetic associations in order to be perceived as useful, but drugs that are covered must be sufficiently described and the contents kept updated. In our opinion, such algorithms should be maintained by local clinicians, geneticists, and pharmacologists, and preferably be included in local clinical protocols.

CONCLUSION

Despite the narrow scope of the prototype, the physicians’ reactions to automatic interpretation of selected genetic variants were very encouraging. Detailed explanations and references are necessary, but should not take center stage in such expert system functionality.

Acknowledgements

We would like to thank Professor Dag Eirik Undlien and Professor Margunn Aanstiid for their guidance in this project, and Professor Aksel Foss and Professor Anna Varberg Reisæter and the physicians at the Department of Transplantation Surgery for their assistance and commitment.

Contributors

HL and TG designed the study protocol and performed initial workflow investigations. HL, TG, SaB and StB produced the overall design of the prototype, and SaB and StB designed and documented the CYP3A5 algorithm and wrote the text templates for the result reports. HL produced the detailed design and programmed the prototype. SaB and StB performed application testing and verification of content. HL performed the investigation and analyzed the recordings and web survey results. HL drafted the article, and TB, SaB and StB revised it critically. All authors were involved in the final approval of the article.

Funding

The research was funded as part of the genAP project by the VERDIKT research program of the Norwegian Research Council. The study sponsor evaluated the general approach of the genAP project, but had no involvement in the study design, the collection, analysis, or interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Competing interests

None.

Ethics approval

Project Steering Group.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

On behalf of the authors, I declare that the data of this study may be shared freely.

REFERENCES
