German Medical Data Sciences: Visions and Bridges
R. Röhrig et al. (Eds.)

doi:10.3233/978-1-61499-808-2-207

Evaluation of an Interactive Visualization Tool for the Interpretation of Pediatric Laboratory Test Results

Johannes HIRSCHMANN [a], Brita SEDLMAYR [a], Jakob ZIERK [b], Manfred RAUH [b], Markus METZLER [b], Hans-Ulrich PROKOSCH [a] and Dennis TODDENROTH [a,1]
[a] *Medical Informatics, Univ. of Erlangen-Nürnberg, Erlangen, Germany*
[b] *Department of Pediatrics and Adolescent Medicine, University Hospital Erlangen*

Abstract. The physiological age-related development of pediatric laboratory results interferes with pathological derangements, which can complicate the interpretation of test results. Recently proposed continuous reference intervals (RIs) promise to be beneficial, although their clinical use may depend on graphical presentations. To estimate the clinical utility of continuous RIs, we developed and evaluated an interactive visualization tool, and examined the differentiation of hemoglobinopathies that is attainable based on the underlying innovative RI model. The implemented web application allows users to easily enter laboratory test results, and displays various visualizations in conjunction with the corresponding RIs, such as charts and personalized Z-scores. To evaluate the usability of the visualization tool, we conducted concurrent think-aloud sessions with four physicians, who were prompted to solve a set of typical interpretation tasks, and acquired additional information through a questionnaire including the System Usability Scale (SUS). We used 85 de-identified clinical cases for an exemplified assessment of how well model-based interpretations of blood count parameters reproduced previously diagnosed hemoglobinopathies. Usability tests as well as questionnaire responses indicated that the developed tool was well received by the physicians. Results from the think-aloud evaluation revealed only minor problems and the tool reached an average SUS score of 86.9, suggesting good usability. Hemoglobinopathy discrimination depended on the considered subtype, although the overall performance of the novel method rivaled the one of the conventional approach. The interactive visualization of innovative continuous reference intervals demonstrated promising results, which justifies further testing on the path towards clinical routine.

Keywords. Technology assessment, evaluation studies, medical informatics application

1. Introduction

Clinicians commonly interpret laboratory tests in relation to established reference intervals (RIs). These limits serve to differentiate between normal and pathological findings, and to gauge the severity of any abnormal increases or decreases. In healthy children, several laboratory parameters are known to characteristically vary with age,

[1] Corresponding Author, Dr. Dennis Toddenroth, Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany, dennis.toddenroth@fau.de

so pediatric test results are usually interpreted in relation to these physiological dynamics. Pediatricians have a particular interest in carefully analyzing available findings, because obtaining blood samples can be more stressful for younger patients.

Conventionally, RIs for pediatric patients have been calculated by partitioning data from a healthy reference population into age groups, so that the corresponding subset-specific percentiles would form step functions of age. More recently proposed statistical approaches instead derive continuous models of the physiological development, which seem biologically more plausibly. The resulting graphical appearance of these modern RIs resembles the familiar steadiness of established anthropometric growth curves [1]. Visual representations of clinical time series generally promise to support diagnostic considerations, for example by promoting a faster recognition of gradual trends [2, 3].

Even if automated graphics creation may theoretically facilitate medical decision-making, practitioners frequently experience that the involved electronic data processing can also introduce tedious obstacles [4]. For a systematic analyses of such intricacies, research on human-computer-interaction has defined usability as *'the extent to which a product can be used by specified users to achieve specific goals'*, which for software applications can depend on the *'capability […] to be understood, learned [and] operated […] under specific conditions'* [5]. To evaluate usability, concurrent think-aloud (TA) prompts representative users to verbalize their thoughts while accessing the studied system for typical tasks. This instrument requires that user comments and activities are recorded, but may permit inferences about cognitive processes [4].

The diagnostic constellations where laboratory tests can be relevant include the identification of hemoglobinopathies, which are oxygen transport diseases that can result in anemia of variable severity. The analysis of the complete blood count usually constitutes only the first step, and further classification of these conditions involves more specific tests [6]. To explore their potential utility for assessing pediatric hematology test results, we thus analyzed how well these next-generation reference intervals reproduced a set of formerly diagnosed hemoglobinopathies in comparison to conventional discrete RIs, and evaluated the usability of a functional visualization prototype.

2. Methods

The previously published method for inducing novel-type continuous RIs leans on estimating the composition of normal and abnormal values in data from routine care; uninterrupted percentiles are then interpolated from overlapping age strata [1]. Our prototypical tool for interactively visualizing such percentile charts together with user-entered blood count values from individual patients was implemented as a web application based on conventional server-side technical platforms. Frontend interactivity was realized via jQuery, while PHP scripts dynamically generate chart graphics and downloadable multi-page PDF files.

Users can enter the patient's sex, birth date and test results from a set of measurement series. Each series includes up to nine hematological analytes as well as the corresponding measurement dates, which can be easily entered via a specialized calendar widget. Besides the plotted percentiles (see figure 1), age-adjusted Z-scores are dynamically calculated for each value and are embedded in the form of „micro-graphics" next to the input fields. A simple color coding scheme highlights

pathological results. All user-entered data can be encoded as and later recalled from Uniform Resource Locators (URLs). An English version of the visualization tool can be accessed at http://www.pedref.org/.

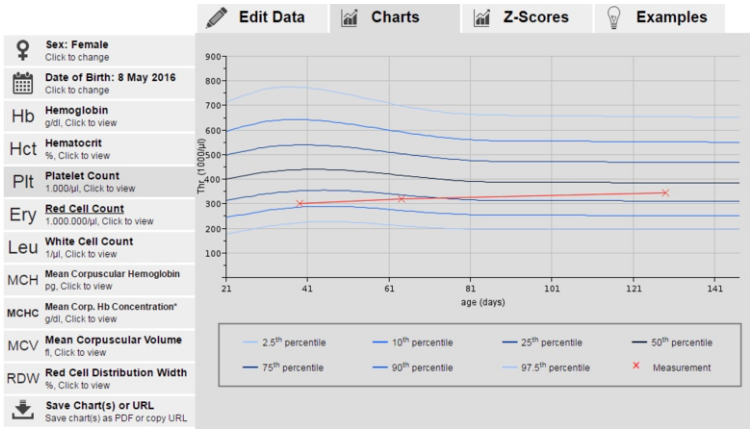

Figure 1. Screenshot of an English version of the visualization tool showing the „Charts" view for the platelet count. Age-dependent percentiles are presented in shades of blue, while current patient measurements are shown in red.

To assess how well the underlying continuous percentile models discriminate between relevant pathology and physiological variation, we considered the exemplary setting of hemoglobinopathy classification based on blood count analysis. The ground truth of our investigation was a de-identified dataset of blood count parameters, taken from initial blood draws of approximately 100 patients including diagnoses. According to the pertinent literature on hemoglobinopathies, we developed simple classification rules, such as *„Is the Hb value below the 2.5th percentile of the standard RI?"* To statistically compare classifier performance, we calculated sensitivity, specificity, as well as receiver operating characteristic curves.

For the usability analysis, we prepared two interpretation tasks of intermediate difficulty; these scenarios required that the test persons access different functions of the visualization tool for interpreting given clinical cases. The test setup consisted of a laptop and a webcam with a microphone. Besides audio and video records, we also captured their onscreen activities to gather task-specific information about possible usability flaws.

Four pediatricians with professional experience ranging between 5 and 14 years participated in separate TA sessions of approximately 30 minutes, held in the pediatrics hospital in Erlangen. During the TA sessions, we tried to keep the level of interference as low as possible, only reminding subjects to think aloud after longer phases of not verbalizing [7]. To gather physician perceptions of the interactive visualizations, we prompted participants to fill out a questionnaire that contained 32 multiple choice items regarding usability, required effort, as well as perceived benefits of the application. From the responses, we calculated system usability scale (SUS) [8] scores.

3. Results

After excluding 15 of the 100 cases due to certain missing blood count values or diagnoses, we analyzed hemoglobinopathy differentiability based on the remaining 85 cases. Inspecting blood count parameters in relation to the RIs revealed that thalassemia diseases might be differentiated from all other observed diagnoses by evaluating the mean corpuscular volume, which seemed to be sharply decreased especially in the case of thalassemia minor. Minor and major forms of thalassemia, as well as homozygosity and heterozygosity of the sickle cell disease, appeared to be distinguishable in a comparable fashion.

In these discrimination scenarios, high classifier performances were achieved. However, differences between employing standard reference intervals, continuous reference intervals and Z-scores were diminishingly small. For example, the resulting area-under-the-curve values regarding the distinction between the 25 cases of thalassemia minor and the remaining 60 diseases were 0.949 (standard RIs), 0.945 (continuous RIs) and 0.912 (Z-scores).

Observations from the TA tests indicated that test subjects were able to complete most tasks with ease. The responses to the questionnaire indicated that our developed application is by and large usable. The items that concerned the data entry options and the Z-score micrographics, for example, were rated very well. However, the evaluation also yielded suggestions for improving certain features, such as connecting it to the hospital information system in order to automatically retrieve patient data via technical interfaces.

Two users had difficulties with the save function, because they were unaware of the German word for „clipboard" (copy-to-clipboard button). This can be seen as a minor issue that could be addressed via various possible solutions. Another suggestion for a potential optimization was given in a comment that implied that the colors of the different percentiles in the charts were not very easy to discriminate.

In total, however, an average SUS score of 86.9 with a standard deviation of 8.5 was obtained, indicating good overall usability. The expert interpretations of the clinical cases often did not match our previously specified diagnoses exactly, although their responses were very close to our expected solutions, and trends were recognized correctly. The additional feedback gathered after the TA sessions stated that the visualization tool could be especially beneficial for assessing more complex clinical cases.

4. Discussion

Our analyses demonstrate that all three models achieved a similar capability to reproduce previously diagnosed hemoglobinopathies. However, a detailed differentiation does not seem to be possible without more advanced tests that involve modern medical apparatus. Furthermore, the sample size was limited, and some patients might even have received blood transfusions before the blood draw, which could have distorted some findings. The good classification performance of all three reference models, on the other hand, indicates that an automated differentiation might be feasible in some scenarios. Further studies of a broader spectrum of hematological data could plausibly reveal additional benefits in other settings.

In order to assess the usability of our visualization tool prototype, we applied the TA protocol, a user-based usability evaluation method. Expert-based methods are seen as more suitable for earlier stages of development [4]. The standardized SUS score that we had measured was considerably above SUS scores that are often referenced in the literature about healthcare sector products [9; 10]. We interpret our obtained results as rather satisfying, and although test subjects were novice users, the visualization tool was well received. The additional feedback can be seen as a sign of interest in the modern visualization of continuous RI. While the usability assessment was done under controlled circumstances, investigations in clinical settings in various institutions could be a reasonable next step to possibly confirm and widen the findings from our experiments.

5. Conclusion

A highly usable tool for visualizing innovative continuous RI was developed. The promising results call for further research, especially towards the impact of integrating such a tool into daily routines.

6. Acknowledgement

Some of the results that are reported here have been previously submitted as part of JH's Master's thesis at Erlangen University. The physicians from the Department of Pediatrics and Adolescent Medicine of the University Hospital Erlangen are gratefully acknowledged for their precious time and effort in participating in this study.

References

[1] Zierk, J.; Arzideh, F.; Haeckel, R.; Rascher, W.; Rauh, M. & Metzler, M. Indirect determination of pediatric blood count reference intervals Clinical Chemistry and Laboratory Medicine, 2013;51:863–87
[2] Torsvik, T.; Lillebo, B. & Mikkelsen, G. Presentation of clinical laboratory results: an experimental comparison of four visualization techniques J. Am. Med. Inform. Assoc., 2013;20:325 - 331
[3] Bauer, D.; Guerlain, S. & Brown, J. The design and evaluation of a graphical display for laboratory data J Am Med Inform Assoc., 2010;17:41 6- 424
[4] Jaspers, M. A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence Int J Med Informatics, 2009;78:340–353
[5] Fernandez, A.; Insfrana, E. & Abrahãoa, S. Usability Evaluation Methods for the Web: A Systematic Mapping Study Information And Software Technology, 2011;53:789-817
[6] Kutlar, F. Diagnostic Approach to Hemoglobinopathies Hemoglobin, 2007; 31:243-250
[7] Boren, T. & Ramey, J. Thinking Aloud: Reconciling Theory and Practice IEEE Transactions on Professional Communication, 2000;43:261-278
[8] Brooke, J. SUS - A quick and dirty usability scale, Digital Equipment Co Ltd., Reading, United Kingdom, 1986
[9] Kortum, P. & Peres, C. S. Evaluation of Home Health Care Devices: Remote Usability Assessment JMIR Human Factors, 2015, 2, e10
[10] Brooke, J. SUS: A Retrospective Journal of Usability Studies, 2013;8:29-40