Pro-Active Detection of Potentially Wrong Diagnoses Due to Substantial Changes of Laboratory Measurements

Motaas ABU BAKER[a], Fabienne CALDERARA[a], Jürgen HOLM[a] and
Murat SARIYAR[a,1]

[a] *Bern University of Appl. Sciences, Dept. Medical Informatics, Switzerland*
[b] *Center for laboratory medicine (ZLM) St Gallen, Switzerland*

Abstract. For guiding decisions on medical diagnoses and diagnoses, it is crucial to receive valid laboratory test results. However, such results can be implausible for the physician, even if the measurements are within the range of known reference values. There are technical sources of implausible results that are related to the laboratory environment, which are frequently not detected through usual measures for ensuring technical validity. Here, we describe the development of a quality assurance tool that tackles this problem and replaces the current manual statistical analyses at the Center for Laboratory Medicine in St Gallen (ZLM). Further analysis of the factors responsible for shifts in laboratory test results requires to collect and analyze data related to reagents as well as calibration or reference probes. Due to a lack of standard operating procedures in many laboratories with respect to these processes, this remains one of the big challenges.

Keywords. Clinical validation, laboratory tests, medical diagnosis, data quality

1. Introduction

Physicians rely on valid results of laboratory tests, when making decisions on diagnoses and therapies, especially in times of COVID testing. However, results can be implausible for the physician, even if the measurements are within the range of known reference values. In a previous work [1], we have dealt with the analysis of factors for erroneous laboratory test results outside the laboratory context. This kind of analysis is referred to as *medical validation* and deals with non-technical sources for implausible values, e.g., a value of 6.0 pmol/L for free triiodothyronine (fT3) resulting from a contaminated blood test of a patient without hyperthyroidism. In addition to that, there are also technical sources of implausible results that are related to the laboratory environment, which are frequently not detected through usual measures for ensuring technical validity [2,3]. For example, reagent lot changes (e.g., of antibodies) can be accompanied by a change of the calibration material that together led to a calibration curve within the same ranges as before, even though the related results for the patients will be shifted. In other words, even if measured values in a technical quality control procedure are within the range of

---

[1] Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

reference values, implausible test results can occur in practice. Physicians should not be confronted with such test results.

Here, we describe the development of a quality assurance tool with R Shiny that replaces the current manual statistical analyses at the Center for Laboratory Medicine in St Gallen (ZLM). To avoid complaints of physicians and reduce the number of suboptimal decisions due to biased laboratory test results, the ZLM wanted to establish a new statistical quality assurance tool to be used in the laboratory only. Two key requirements were, (i) that a user with limited IT knowhow should be able to use the tool, and (ii) that automated statistical evaluations can be triggered. In the end, the current manual statistical analyses are replaced using a standard frontend to perform the necessary validation work. Further, interactive, and annotated figures, which can be produced on the fly, facilitate verification of the computed statistical values.

2. Methods

First, we derived user requirements of the ZLM through user stories. A user story describes the features of a software system from the perspective of the user [4], for example: User KJ (who) wants a moving average graph for all fT3 test results from 2019-01-01 to 2019-12-31 (what) in order detect significant shifts in the measurements (goal). After that, we developed a mock-up to ensure to have captured all requirements. Finally, we developed a web-based prototype based on the R shiny framework (see [5]) and anonymized data related to laboratory devices stemming from laboratory information system of the ZLM. The decision for R as programming environment was motivated by the large number of packages for statistical analysis and the sophisticated visualization capabilities.

In addition to that, we conducted a literature search using the PubMed database. We were only interested in studies that aimed at technical validation and the calibration process by means of statistical quality assurance. Following combination of keywords were used:

- "technical validation" AND "calibration" AND "quality assurance"
- "technical validation" AND "medical validation"
- "technical validation" AND "clinical validation"
- "technical validation" AND "clinical validation" AND "statistics"
- "technical validation" AND "statistical quality control"
- "technical validation" AND "calibration process"
- "technical validation" AND "statistical"

3. Results

In the literature research, we did not find any study that addresses with the issue of proactively detecting substantial changes of laboratory test results caused by technical processes that are not detected by common technical validation procedures. The focus is either on technical validation or on medical validation, neglecting the impacts of laboratory device changes on laboratory parameter measurements not detected in standard validation procedures are not tackled. We assume two central reasons for this unexpected result: (i) reliance on the technical validation, once it is established and (ii)

a high number of unreported cases in practice, especially when the shifts in the test results are rather minor (unknown unknowns cannot be investigated). With respect to the development of our tool, the outcome of the literature research meant, that we had no further inputs other than the requirements of our stakeholder. To facilitate generalizability of our prototype, we concentrated on statistical analysis that are easy to interpret and allowed many options for going into details, e.g., by selecting date ranges.

Our prototype has three main tabs: monitoring (automatic evaluation), manual evaluation and settings. Per default, the monitoring tab is selected when accessing the web application. For all laboratory parameters, daily, weekly, and monthly changes of the median measurement values are computed and visualized with a density plot (see Figure 1). The median value changes are computed as (representing relative changes)

$$\frac{median_t - median_{t-1}}{median_{t-1}},$$

with *t* ranging over days, weeks, or months. An automated generated email is sent out by the prototype, when the absolute values exceed a certain threshold. Furthermore, reactive moving median curves are generated in a subtab of the monitoring tab, allowing to inspect the development of measurement values. Within the manual evaluation tab, a broader range of statistical analyses becomes available, e.g., hierarchical clustering and statistical tests, allowing the user to investigate possible reasons for certain changes.
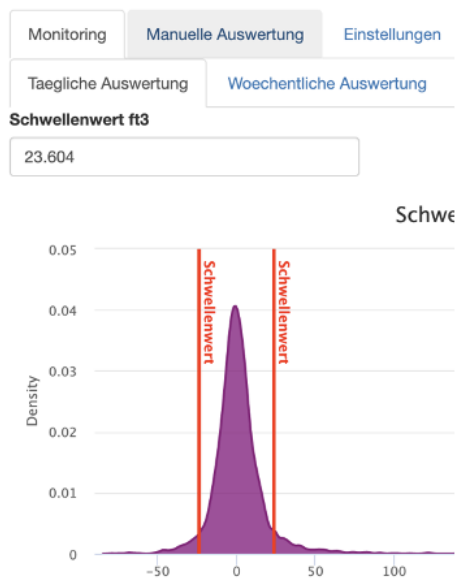

Figure 1. Density plot of relative changes of the median (y) for fT3 values with corresponding thresholds (Schwellenwert), for which only the right one (here the 90% percentile) is relevant in our datasets.

We discovered on our data, that the most important insights were given by the moving median curves of the measurement values summarized per week (daily data were too granular and monthly data too coarse). The user of our prototype is enabled to change the smoothing parameter to determine whether to focus on short- or long-time effects (trends). In Figure 1, the relevant shift of the fT3 measurements in December 2015 is

clearly discernible. Due to the interactivity of the graphs, by using the R-package plotly, the user gets annotation of data points and can zoom in and out. The reason for the peak in January 2015 is not known.
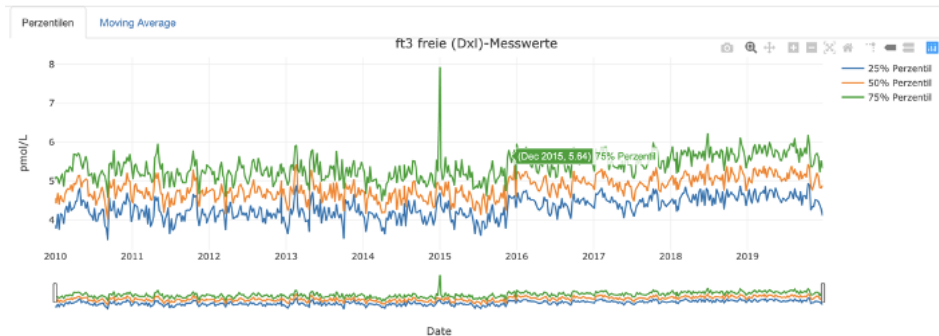

Figure 2. Moving median curve for the fT3 values in a reactive plotly graph.

4. Discussion

Proactively detecting substantial changes of laboratory test results caused by technical processes that are not detected by common technical validation procedures is still uncommon, as there is often reliance on the technical validation. For the analysis of factors outside standards technical validation responsible for shifts in laboratory test results, it is crucial to include valid data related to reagents, calibration, or reference probes in addition to time series of test result values. Due to a lack of standard operating procedures in laboratories, this remains one of the big challenges.

An implication of substantial changes of laboratory test results is the adjustment of the corresponding reference intervals. At the ZLM, there is a standardized procedure for setting a reference interval. The prototype will increase its usefulness, when it integrates and enriches this procedure, allowing to address all issues around the shifts of measurement values. Following insight should be considered in this context [6]: "Analytical factors and pre-analytical factors need to be considered, along with partitioning on the basis of sex or age, either as part of a reference interval study or when interpreting other studies."

References

[1]   Degen GA, Günther V, Holm J, Bürkle T, Sariyar M. Using Business Intelligence Tools to Support Medical Validation of Laboratory Tests. *Stud Health Technol Inform*. 2020; **270**:494–498.
[2]   Jennings L, Van Deerlin VM, Gulley ML, College of American Pathologists Molecular Pathology Resource Committee. Recommended principles and practices for validating clinical molecular pathology tests. *Arch Pathol Lab Med*. 2009; **133**:743–755.
[3]   Kay JF. Combining ISO/IEC 17025:2005 and European Commission Decision 2002/657 audit requirements: a practical way forward. *Drug Test Anal 4 Suppl*. 2012; **1**:25–27.
[4]   Kannan V, Basit MA, Bajaj P, et al. User stories as lightweight requirements for agile clinical decision support development. *J Am Med Inform Assoc*. 2019; **26**:1344–1354.
[5]   Kraus M, Mathew Stephen M, Schapranow MP. Eatomics: Shiny Exploration of Quantitative Proteomics Data. *J Proteome Res*. 2019; **20**:1070–1078.
[6]   Jones G, Barker A. Reference Intervals. *Clin Biochem Rev.* 2008; **29**: S93–S97.