Clinica Chimica Acta 445 (2015) 12–18

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Monitoring laboratory data across manufacturers and laboratories—A prerequisite to make "Big Data" work

Kenneth Goossens [a], Katleen Van Uytfanghe [a], Patrick J. Twomey [b], Linda M. Thienpont [a,*], Participating Laboratories [a,b]

[a] Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ghent, Belgium
[b] Department of Clinical Chemistry, St Vincent's University Hospital, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 3 March 2015
Accepted 4 March 2015
Available online 12 March 2015

Keywords:
Quality indicator
Median
Moving median
Stability of performance
Shifts
Drifts

ABSTRACT

Background: "The Percentiler" project provides quasi real-time access to patient medians across laboratories and manufacturers. This data can serve as "clearinghouse" for electronic health record applications, e.g., use of laboratory data for global health-care research.
Methods: Participants send their daily outpatient medians to the Percentiler application. After 6 to 8 weeks, the laboratory receives its login information, which gives access to the user interface. Data is assessed by peer group, i.e., 10 or more laboratories using the same test system. Participation is free of charge.
Results: Participation is global with, to date, >120 laboratories and >250 instruments. Up to now, several reports have been produced that address i) the general features of the project, ii) peer group observations; iii) synergisms between "The Percentiler" and dedicated external quality assessment surveys. Reasons for long-term instability and bias (calibration- or lot-effects) have been observed for the individual laboratory and manufacturers.
Conclusions: "The Percentiler" project has the potential to build a continuous, global evidence base on in vitro diagnostic test comparability and stability. As such, it may be beneficial for all stakeholders and, in particular, the patient. The medical laboratory is empowered for contributing to the development, implementation, and management of global health-care policies.

1. Introduction

Laboratory data has the potential for substantially aiding the development, implementation, and management of public healthcare policies. It can create public awareness of the importance of maintaining a healthy lifestyle as well as reacting early to signals of health problems. As such, it can indirectly contribute to reduce the burden of healthcare expenses. The drive to focus on improved exploitation of laboratory data typically comes from financial pressures, such as the steady increase in health-care expenses in the US during the last 20 years. Such expenses now represent 17.6% of the gross domestic product and nearly $600 billion more than the expected benchmark for a nation of the size and wealth of the US [1]. An additional impetus to transform the laboratory landscape comes from the information technology (IT) revolution, offering, among others, the opportunity to create reliable and accessible "Big Data" [2,3]. Nevertheless, the "big bang" for the active role of IT in healthcare policy came in the US from legislation "The Health Information Technology for Economic and Clinical Health (HITECH) Act" [4] and the push by the government to adopt electronic health records (EHRs) [5–8]. The research firm Frost & Sullivan predicts that use of advanced health data analytics solutions in hospitals will increase to 50% adoption by 2016 [9]. This may create enormous business opportunities, for example, the Washington Post reported the inclusion of "as much as $36.5 billion in spending to create a nationwide network of electronic health records" [10]. However, big spending should be justified by big savings. Indeed, according to a report from McKinsey & Company, the largest managed care organization in the US (Kaiser Permanente), reported that their "Big Data" strategy has saved the organization $1 billion in reduced office visits and lab testing [1].

If data is the new gold, then access to data is going to be key to insights [2]; however the expertise of the laboratory is also key to ensure the reliability of the data as well as its safe and efficient use. While promises are sky-high, EHR is not without risks, especially in the start-up phase. In this regard "The November 2011 Institute of Medicine report, Health IT and Patient Safety: Building Safer Systems for Better Care" noted that the lack of empirical data on the nature and prevalence of EHR system-related adverse patient events makes it challenging to

Abbreviations: ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C-reactive protein; EHR, electronic health record; GGT, γ-glutamyl transferase; IT, information technology; LDH, lactate dehydrogenase; LIS, laboratory information system; MCs, master comparisons.

* Corresponding author at: Laboratory for Analytical Chemistry, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, B-9000 Ghent, Belgium. Tel.: +32 9 264 81 04; fax: +32 9 264 81 98.
E-mail address: linda.thienpont@ugent.be (L.M. Thienpont).

http://dx.doi.org/10.1016/j.cca.2015.03.003
0009-8981/

determine the extent of the associated risks to patient safety [11]. If one takes this note seriously, the question is how to minimize risks? Can a small non-US based group play a role in this highly innovative, competitive, multi-billion dollar environment? We believe yes, because we are convinced that efforts to take care of the quality of the input data will improve the quality of the output. The utility of the EHR is often overlooked as a result of laboratory diagnostics which sometimes give significantly different values for the same patient sample, even for the simple, high-volume clinical chemistry tests [12–14].

In an attempt to illustrate this limitation and more importantly to do something about it, we describe "The Percentiler" project, which is part of our overarching "Empower" project introduced elsewhere [15]. In essence, it provides quasi real-time access to patient medians across laboratories and manufacturers. This data can serve as a "clearinghouse" for potential future EHR applications, such as the retrieval of laboratory data for epidemiological or toxicological research on national or global scale, long-term follow-up of chronic diseases, or linking laboratory data to mortality risk [16,17].

2. Materials and methods

2.1. Participants and participation process

Participating laboratories are globally distributed. They range from medium-sized to big hospital laboratories, but also include private laboratories (for the current list of participants, see www.stt-consulting.com, Empower tab). When a laboratory declares its intention to join, we provide it with the information about the IT requirements for sending data, together with a request for a test e-mail. One of our project team controls the test-mail, maps the data and verifies error-free transmission into our database. If successful, we continue this verification for a while before giving the definitive Percentiler e-mail address. Subsequently, data transfer either occurs automatically and on a daily basis (depending on the Laboratory Information System (LIS)) or is done in manual batches. After sending data for 6 to 8 weeks, the participating laboratory receives its login information, which gives access to the graphical presentation of its data via a user interface. Data is assessed by peer group: typically 10 or more laboratories using the same test system. Participation is free of charge. Furthermore, all LIS solutions for automatic median calculation and data transfer are provided at no or minimal cost and without running costs.

2.2. Data

We collect instrument-specific daily medians calculated from outpatient results of 20 commonly measured analytes in serum or plasma: albumin, alanine aminotransferase (ALT), alkaline phosphatase (ALP), aspartate aminotransferase (AST), calcium, chloride, C-reactive protein (CRP), creatinine, γ-glutamyl transferase (GGT), glucose, inorganic phosphorus (phosphate), lactate dehydrogenase (LDH), magnesium, potassium, sodium, total-bilirubin, total-cholesterol, total-protein, urea, and uric acid (urate).

2.3. Data coding and transfer to a database

Data coding comprises 7 attributes each separated by "semicolon": laboratory identification (Lab ID); date (e.g., 02/01/2014); instrument identification (Instr ID); code for outpatients (e.g., OUT); test name (e.g., CA for calcium); test unit (e.g., mmol/L); median (e.g., 2.35). The laboratories can retrieve these attributes directly from the LIS and adopt the used mnemonics. The only requirement is for the laboratories to organize the data in a table according to the format below:

```
Lab ID;02/01/2014;Instr ID;OUT;CA;mmol/L;2.35
Lab ID;02/01/2014;Instr ID;OUT;NA;mmol/L;141
Lab ID;02/01/2014;Instr ID;OUT;CL;mmol/L;102.5
```

Data must be sent electronically to our project-specific e-mail address either as i) an e-mail embedded table, ii) an e-mail attached EXCEL-file; or iii) an e-mail attached text-file. Retrieval of data and electronic exporting is done either automatically (by features available in a specific LIS), or manually. Automatic solutions send the data daily, while manual solutions operate in a batch fashion with the data manually extracted weekly or monthly and manually sent by e-mail.

2.4. Database

The software for data downloading from the e-mail, transfer into a MySQL database, and the development of "The Percentiler" application and user interface was programmed by Bruno Neckebroek (Zwijnaarde, Belgium). Data from the individual laboratories are "mapped" by the STT/UGent project team to common analyte names, units, and instrument names and other technical details.

2.5. Data analysis/user interface

The database is fully accessible to the STT/UGent project team, who investigate laboratory and peer group data for bias and trends. Critical observations are communicated in the first instance to the laboratories concerned. They are also shared with instrument vendors, and regularly, with the whole group of participants. It is important to note that the identity of the laboratory is not disclosed to a third party under any circumstances. The user interface (accessed via a specific login at https://thepercentiler.be) only gives the laboratories access to their own data (login with username and password). Investigation of data is possible on-line. However, if detailed off-line analysis is preferred, the data can be downloaded into Excel®. Other functionality in the user interface allows a downloadable chart of the moving median in time (laboratory and peer), and a table with summary statistics (bias, robust CV) for each analyte. The selection possibilities include i) n for calculation of the moving median ($n = 5, 8, 16$); ii) time window; and iii) inclusion/exclusion of weekends. When participants report medians for two or more instruments, an instrument-specific color code is used in the charts. The assessment of the stability of laboratory testing is done against desirable bias limits from biological variation, at least for the analytes for which state-of-the-art performance allows this [18]. However, the maximum bias limit is set to ~10%. The limits are visualized in the charts by a gray zone, and violations >1 week are considered significant. For more detailed information, the reader is referred to the demo version of "The Percentiler" (https://thepercentiler.be, login: demolab, password: demo1234).

2.6. Partners

The Royal Belgian Society of Clinical Chemistry scientifically supports the project. The assistance from several LIS vendors in providing solutions for automatic data calculation, retrieval, and electronic sending greatly contributed to the practical realization of the project. Further support is received from the Belgian representatives of the main in vitro diagnostic manufacturers (see www.stt-consulting.com, Empower tab, for LIS and manufacturer information).

3. Results

3.1. Participation and reporting

Currently (December 2014), 124 laboratories participate with ~250 instruments, distributed over the following peer groups: Advia ($n = 8$); Architect (19); AU (13); Cobas (153); Integra (3); Modular (11); Synchron (11); Vista (6); and Vitros (26). Participation is global (see Fig. 1); however, most of the current participants come from Belgium.