Clinica Chimica Acta xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Clinica Chimica Acta

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

Monitoring the stability of the standardization status of FT4 and TSH assays by use of daily outpatient medians and flagging frequencies

Linde A.C. De Grande [a], Kenneth Goossens [a], Katleen Van Uytfanghe [b], Barnali Das [c], Finlay MacKenzie [d], Maria-Magdalena Patru [e], Linda M. Thienpont [a,*], for the IFCC Committee for Standardization of Thyroid Function Tests (C-STFT):

[a] Department of Pharmaceutical Analysis, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, 9000 Ghent, Belgium
[b] Ref4U, Laboratory of Toxicology, Faculty of Pharmaceutical Sciences, Ghent University, Ottergemsesteenweg 460, 9000 Ghent, Belgium
[c] Biochemistry and Immunology Laboratory, Kokilaben Dhirubhai Ambani Hospital and Medical Research Institute, Mumbai, India
[d] Birmingham Quality/UK NEQAS, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK
[e] Ortho-Clinical Diagnostics, Inc., Rochester, NY, USA

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form 21 March 2016
Accepted 26 April 2016
Available online xxxx

Keywords:
Committee for Standardization of Thyroid Function Tests
Median
Outpatient
International Federation of Clinical Chemistry and Laboratory Medicine
Quality indicator
Population variation

ABSTRACT

Clinicians diagnose thyroid dysfunction based on TSH and FT4 testing. However, the current lack of comparability between assays limits the optimal use of laboratory data. The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) gave a mandate to the Committee for Standardization of Thyroid Function Tests (C-STFT) to resolve this limitation by standardization. Recently, the Committee members and their partners felt ready to set the step towards the technical recalibration. However, before implementation, they were furthered by the Food and Drugs Administration (FDA) to develop a tool to assess the sustainability of the new calibration basis. C-STFT began to use 2 online applications, i.e., the "Percentiler" and "Flagger", with the intention to assess their utility for this purpose. The tools monitor the course of instrument-specific moving medians of outpatient results (Percentiler) and flagging rates (Flagger) from data of individual laboratories grouped by instrument/assay peer. They additionally document the mid- to long-term medians, hence, are quality indicators of stability of performance of both laboratories and peers/assays. Here, the first experiences built up in the pre-standardization phase are reported. They suggest the suitability of both applications to document the sustainability of the calibration basis in the post-standardization phase.

1. Introduction

Given the prevalence and severity of different forms of thyroid disease, the yearly number of tests performed worldwide is huge [1–4]. Clinicians mainly rely on the analysis of thyroid-stimulating hormone (TSH) and free thyroxine (FT4) for the diagnosis of thyroid dysfunction and patient follow-up. The frequency of thyroid function testing translates in an enormous impact of the disease on the healthcare system. In this regard, it is generally recognized that, to reduce the expenses for healthcare from laboratory analysis, comparability of measurement data over time, location and across assays would be utmost beneficial. Indeed, once this is achieved, laboratory data can meet modern clinical and public health needs, such as the definition and use of common reference intervals/clinical decision limits, development of evidence-based clinical practice guidelines for application of consistent standards of medical care, translation of research into patient care and disease prevention activities, inclusion of laboratory data in electronic patient records, etc. [5]. However, to accomplish this, in depth transformation of the current laboratory landscape in general but for thyroid function testing in particular is required. Indeed, the problem of observed between-assay discrepancies needs to be resolved [6,7].

The International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) decided to pay attention to these needs. In 2005, the Scientific Division formed the Working Group for Standardization of Thyroid Function Tests (WG-STFT) with the mission statement to document the standardization status and intrinsic quality of current thyroid hormone immunoassays. The focus of the activities should be on TSH and FT4 testing, and where necessary, on improving the standardization status [6,7]. In 2012, the WG was transformed into a Committee (C-STFT) to broaden the scope of stakeholders [8].

The achievements of the WG up to now are described elsewhere [6,7,9–16]. They comprise developing reference measurement systems for standardization of FT4 and harmonization of TSH, demonstrating the

Non-standard abbreviations: C-STFT, Committee for Standardization of Thyroid Function Tests; FDA, Food and Drug Administration; FT4, free thyroxine; IFCC, International Federation of Clinical Chemistry and Laboratory Medicine; IVD, in-vitro diagnostic; LIS, Laboratory Information System; TSH, thyroid-stimulating hormone; WG-STFT, Working Group for Standardization of Thyroid Function Tests.

* Corresponding author.
E-mail address: linda.thienpont@ugent.be (L.M. Thienpont).

http://dx.doi.org/10.1016/j.cca.2016.04.032
0009-8981/

feasibility of their use as a uniform calibration basis for commercial in-vitro diagnostic (IVD) thyroid function tests, and designing a step-up approach based on several dedicated method comparison studies to allow new manufacturers to join the standardization activities. Recently the C-STFT set the step to the technical recalibration process of FT4 and TSH assays by a method comparison with clinically relevant panels of samples (results currently under investigation). Although in theory this process completes the establishment of a uniform calibration basis of the assays – at least for diagnosis and follow-up of uncomplicated hypo- and hyperthyroidism – immediate implementation is not possible but needs careful preparation. One of the actions currently undertaken in this regard was that the Committee – comprising laboratory professionals and IVD manufacturers – visited the Food and Drug Administration (FDA) authorities. They presented the technical approach, discussed its acceptability and the plans before implementation. The Committee got a positive response from the FDA, who particularly welcomed the plan to establish a dialogue basis with an as broad spectrum of stakeholders as possible, and investigate with them the benefits but also the risks associated with implementing the standardized/harmonized IVD assays. The benefit-risk analysis recently has been initiated, among others, at the level of medical laboratories (internally by consultation of delegates designated by the IFCC Member Societies) and clinicians/patients [e.g., 17]. In addition, the FDA furthered the Committee to document – preferably at the level of real patient results – the sustainability of the post-standardization calibration status of the participating IVD assays. The Committee got by courtesy of STT-Consulting and the Chair (currently Thienpont & Stöckl Wissenschaftliches Consulting) access to 2 new quality management tools to assess whether they could serve the above purpose implied by the FDA. Note that as described elsewhere the tools are part of the overarching "Empower project" [18–21]. One tool, called the "Percentiler" monitors daily outpatient medians to reflect the stability/variation of performance at the level of the individual laboratory and its peer group. Its potential to build a global evidence base on IVD test stability across laboratories and peers/manufacturers has been shown before from application for clinical chemistry analytes. The second tool, the "Flagger", monitors flagging of results against reference intervals or decision limits used in the individual laboratory, but also at the peer group level. It is complementary to the "Percentiler" in that it directly translates the effect of analytical quality/(in)stability on flagging as surrogate medical decision making [22]. In view of this potential the C-STFT decided to start using the Percentiler and Flagger (in cooperation with the Empower team) in the framework of its standardization activities. One important matter of concern that needed investigation was whether the tools would similarly be useful to monitor the stability of FT4 and TSH assays as they are for clinical chemistry ones, particularly, because it could be anticipated that the reported median and flagging rate values would be based on a substantially lower number of results per day. The Empower team invited laboratories, already using the applications for clinical chemistry analytes, to extend their participation to FT4 and TSH. For obvious reasons, the focus was on laboratories using the IVD test systems/assays involved in the C-STFT standardization/harmonization project. The C-STFT's intention was to explore the utility of both tools in the pre-standardization phase, and if positive, to fully exploit them in the post-standardization phase for the purpose implied by the FDA. Here, we report on behalf of C-STFT on the experience built up in the pilot study.

2. Material and methods

The way the data are collected in both applications has been described in detail elsewhere [19–21]. Participation is free of charge. In brief, laboratories calculate – preferably by an automatic function in their Laboratory Information System (LIS) or, if not available, manually – instrument-specific daily medians (preferably) from outpatient results. The data are automatically sent by e-mail on a daily basis or batch wise to the Percentiler's and Flagger's MySQL database. For the Flagger application laboratories also report the daily flagging rate in percentage of the total number of generated results. Whereas the Empower team can investigate the complete database at the individual laboratory and peer group level, the participating laboratory only has access to its own data via a user interface (to access via a specific login and password at https://thepercentiler.be and https://theflagger.be, respectively). These interfaces have several functionalities, such as downloadable charts of the laboratory's instrument-specific moving medians of patient results (Percentiler) and flagging rates (Flagger) in time, as well as a table with summary statistics (bias, robust CV). The moving median charts also show the mid- to long-term medians of the laboratory and its peer group. In the Percentiler application the respective numerical values are documented in the statistics table, where also a target median is given (see below). The laboratory bias is compared to the peer group and target median. The deviation (in %) from the target is evaluated against desirable bias limits from biological variation, i.e., 3.3% for serum FT4 and 7.8% for serum TSH, respectively [23]. These limits are visualized in the charts by a gray zone around the laboratory's mid- to long-term median to reflect the stability of performance. Violations of the limits that last > week are considered significant. With regard to the aforementioned target median used to assess the bias of the individual laboratory and peer group medians, currently the all-laboratories' median is utilized. In the pre-standardization phase the overall median for FT4 is 15 pmol/L with $\pm$0.5 pmol/L as limits, for TSH 1.56 mIU/L $\pm$ 0.12 mIU/L, respectively. In the Flagger application a certain relative percentage around the long-term median (with an absolute minimum of 1%) is used as the limit, which should not be violated. For FT4 and TSH the limit is preliminary set to 30% [22].

3. Results

In the Percentiler application, currently (March 2016) 76 laboratories participate with 158 instruments, while in the Flagger, 33 laboratories supply data from 44 instruments. The number of laboratories and test systems per manufacturer are listed in Table 1, including the average participation time per peer group. This should give an indication of the number of data points accounted for in this pilot study (1 data point (1 median value) per assay per instrument per day is received). For this study, we distinguished in the Percentiler between 5 peer groups, i.e., Roche Cobas, Siemens Centaur, Abbott Architect, Beckman Synchron, and OCD Vitros, whereas in the Flagger, only the Roche Cobas peer group is currently sufficiently substantiated. Therefore, the data given for this application are very preliminary.

We calculated from the patient data in the Percentiler the respective peer group medians for both FT4 and TSH. For FT4 the peer group medians ranged in the pre-standardization/harmonization phase from ~11.7 to 16 pmol/L, for TSH from ~1.2 to 1.7 mIU/L, respectively. In Fig. 1, the match of the peer group medians in this pilot study with those from the previous Phase I method comparison studies is shown [6,7].

Although the time period of participation is still short for the majority of assays (on average 11 months), most laboratories generally showed a stable performance for both analytes. However, in some individual cases we observed greater variation in performance (drifting or shifting medians), occurrence of a transient bias, between-instrument differences within a laboratory, etc. A few representative examples are given: in Fig. 2A a laboratory is documented with an acceptable analytical stability for TSH analysis on all instruments; indeed all moving medians are nicely between the stability limits and concordant with the peer group median; in contrast, Fig. 2B shows a laboratory with highly variable FT4 moving medians for all instruments it uses; in Fig. 2C the concerned laboratory has clear shifts in its FT4 performance (note the moving medians exceeding the stability limits), while the laboratory shown in Fig. 2D performs for TSH with a substantial difference between the 2 instruments it uses.

Please cite this article as: L.A.C. De Grande, et al., Monitoring the stability of the standardization status of FT4 and TSH assays by use of daily outpatient medians and flagging frequ..., Clin Chim Acta (2016), http://dx.doi.org/10.1016/j.cca.2016.04.032