Editorial

Mario Plebani and Luca Giovanella

Reflex TSH strategy: the good, the bad and the ugly

https://doi.org/10.1515/cclm-2019-0625

The laboratory diagnosis of thyroid dysfunction relies on the measurement of circulating concentrations of thyrotropin (TSH), free thyroxine (fT4), and, in some cases, free triiodothyronine (fT3). TSH and fT4 have a complex, nonlinear relationship, such that small changes in fT4 result in relatively large changes in TSH [1]. Even if some rare exceptions exist (i.e. central hypothyroidism, resistance to thyroid hormones, TSH-secreting pituitary adenoma, treated hyperthyroidism and nonthyroidal illness), TSH measurement is a sensitive screening test for thyroid dysfunction and guidelines from the American Thyroid Association [2], the American Association of Clinical Endocrinologists [3] and the National Academy of Clinical Biochemistry [4] have all endorsed its measurement as the best first-line strategy for detecting thyroid dysfunction in most clinical settings. Then, to reduce the need for fT4 testing without compromising the detection of overt thyroid dysfunction fT4 may be added to existing requests, either automatically on the basis of algorithms (i.e. reflex testing) or by laboratory professionals (i.e. reflective testing). These strategies proved to be clinically appropriated and cost-effective in first-line assessment of thyroid function, but few data are reported in the literature on the importance of establishing appropriate cutoffs to avoid the risk of undetected hyperthyroidism and/or hypothyroidisms. TSH limits to launch reflex or reflexive fT4 measurements has been the subject of controversy and, in particular, a seminal paper by Henze and colleagues [5] evaluated the effect of different TSH cutoffs for reflex testing of fT4. In a clinical cohort of subjects referred for thyroid function testing, the use of wider cutoffs to trigger reflex testing resulted in a substantial reduction in the need for fT4 testing compared with when the TSH reference range limits are used. Applying TSH cutoffs of 0.3 and 5.0 mU/L for reflex testing in place of the reference range limits of 0.4 and 4.0 mU/L resulted in a 22% reduction in fT4 tests, whereas cutoffs of 0.2 and 6.0 mU/L reduced the fT4 testing by 34%. In the community cohort, the effect of these cutoffs was smaller with the corresponding reduction in fT4 testing of 3.2% and 4.8%, reflecting the lower prevalence of thyroid dysfunction in the general community. In this issue of the journal, Taher and colleagues [6] report on their experience with TSH reflexive testing implemented at their hospital in 2016. The algorithm reflexed fT4 and fT3 when TSH was low (i.e. <0.40 mIU/L) and fT4 when TSH was high (i.e. >5.50 mIU/L). To assess the appropriateness of TSH reflexive testing cutoffs, 3 years of historical data (n = 87,465) prior to reflex implementation were collected. ROC curves were generated using R (nonparametric trapezoidal approach) to identify optimal low and high TSH cutoffs by using both the Youden index (YI) and the Euclidean distance (ED). Lower and upper values were 0.4 mIU/L (both methods) and 5.67 mIU/L (YI) vs. 5.58 mIU/L, respectively. Thus, altering cutoffs from the current reference range would remove less than 1% of tests confirming the appropriateness of using their preexisting reference range to trigger the reflexive testing algorithm (i.e. 0.40–5.50 mIU/L). Their results diverged from those previously published by Gill and colleagues [7] showing that the need for additional reflex testing for fT4 values could be reduced with minimal clinical effects by widening the normal reference range for TSH. A comparison between different studies is difficult due to different criteria adopted to select patient and control populations (with the related different prevalence of thyroid dysfunctions), different results due to different laboratory instrumentation and reference populations and the influences of heterogeneity in the TSH molecule and interfering substances. Last but not least the clinical context should be considered to design an appropriate approach to the thyroid function testing. In fact, in specialized thyroid centers, laboratory tests are selectively required based on a well-defined clinical problem. Here, sequential case-specific test(s) request(s) is preferred and data are evaluated using clinically-oriented interpretation criteria and frequently discussed in a multidisciplinary setting (i.e. thyroid board). It is important to note, however, that in general clinical practice thyroid tests are mainly required to rule-out thyroid dysfunctions in patients with unspecific symptoms: in this context what is needed is a test able to safely rule-out diseases.

Taken together, the take home messages from these two papers [6, 7] highlight the following points:

a) The reflex testing based on TSH levels is a tool for reducing additional and inappropriate fT4/fT3 tests when screening people for new thyroid disease but is not appropriate for certain patient groups, including those with known established thyroid diseases or suspected pituitary disease. It should be highlighted that central hypothyroidism (CeH) has been defined as "*a neglected thyroid diagnosis*" as its frequency is underestimated [8]. In this respect, the missed diagnosis of CeH represents the most important false-negative result of the "reflex TSH" strategy.
b) TSH cutoffs are strongly affected by the different accuracies of immunoassays, because inter method differences of about 1 mU/L at concentrations of 4–5 mU/L have been reported [9]. According to the current state-of-the-art, and poor harmonization of TSH assays, the lack of interchangeability of laboratory results and cutoffs does not allow the proposal of common cutoffs to be used in the "reflex TSH" strategy. Therefore, each clinical laboratory, unfortunately, has still to establish reliable cutoffs based on the adopted method.
c) In addition, adopted cutoffs may be different on the basis of the specific populations and the clinical context, as in the general population the approach to be used is to adopt the cutoffs allowing the most relevant reduction of fT4 tests, while in other clinical settings a different cutoff should be preferable to avoid missing hypo- and/or hyper-thyroidism diagnoses.
d) The "reflex TSH" strategy is *good* when appropriate cutoffs are applied to the right population and in the right clinical context. It is *bad* when the cutoffs are based on the reference ranges suggested by manufacturers and/or when not-validated cutoffs are used. It should be *ugly*, if applied to the wrong patients, with wrong cutoffs and using poorly validated immunoassays.

In conclusion, the "reflex TSH" strategy should be viewed as an opportunity to improve appropriateness in test requests and for saving unjustified costs for healthcare systems. However, this should be closely linked to the need to avoid missing cases of thyroid disorders. In addition, these papers,reinforce the need, once again, to support initiatives to improve standardization and harmonization in laboratory medicine.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Employment or leadership: None declared.
Honorarium: None declared.

References

1. Koulouri O, Moran C, Halsall D, Chatterjee K, Gurnell M. Pitfalls in the measurement and interpretation of thyroid function tests. Best Pract Res Clin Endocrinol Metab 2013;27:745–62.
2. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association Task Force on Thyroid Hormone Replacement. Thyroid 2014;24:1670–751.
3. Garber J, Cobin R, Gharib H, Hennessey J, Klein I, Mechanick J, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract 2012;18:988–1028.
4. Demers LM, Spencer CA. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. Clin Endocrinol 2003;58:138–40.
5. Henze M, Brown SJ, Hadlow NC, Walsh JP. Rationalizing thyroid function testing: which TSH cutoffs are optimal for testing free T4? J Clin Endocrinol Metab 2017;102:4235–41.
6. Taher J, Brinc D, Gilmour JA, Beriault DR. Validating thyroid-stimulating hormone (TSH) reflexive testing cutpoints in a tertiary care institution. Clin Chem Lab Med 2019;58:e11–3.
7. Gill J, Barakauskas VE, Thomas D, Rodriguez-Capote K, Higgins T, Zhang D, et al. Evaluation of thyroid test utilization through analysis of population-level data. Clin Chem Lab Med 2017;55:1898–906.
8. Beck-Peccoz P, Rodari G, Giavoli C, Lania A. Central hypothyroidism – a neglected thyroid disorder. Nat Rev Endocrinol 2017;13:588–98.
9. Kahapola-Arachchige KM, Hadlow N, Wardrop R, Lim EM, Walsh JP. Age-specific TSH reference ranges have minimal impact on the diagnosis of thyroid dysfunction. Clin Endocrinol 2012;77:773–9.

Corresponding author: Mario Plebani Department of Laboratory Medicine, University-Hospital, 3218 Padova, Italy,
E-mail: mario.plebani@unipd.it.
https://orcid.org/0000-0002-0270-1711
Luca Giovanella: Clinic for Nuclear Medicine and Competence Centre for Thyroid Diseases, Ente Ospedaliero Cantonale, Bellinzona, Switzerland; and Clinic for Nuclear Medicine, University Hospital and University of Zurich, Zurich, Switzerland