

Prof. Dr. Rudolf Hoermann

To The Committee
on Thyroid and Adrenal Diagnosis
Scottish Parliament
petitions@scottish.parliament.uk

Dear Madam and Sir:

I am writing regarding the recent petition on the return to appropriate thyroid and adrenal diagnosis.

The reason for doing so is that I have listened to the public broadcast of the committee hearing on the internet where a recent publication of our group has been mentioned. I have been made aware of the issue by Thyroid UK, but I don't speak for any organisation and express my own private views.

To briefly introduce myself, I am currently semi-retired to the Sunshine Coast, Australia following a long career as a professor of Endocrinology and head of department in a teaching hospital in Germany. I have also some knowledge of the American system during my training with Harvard University early on in my career and I am a member of the American, European and German Endocrine Societies as well as the American and German Thyroid Association. I still have a ongoing interest in the field of endocrinology/ thyroidology including recent publications in peer reviewed journals, one of which has been referred to by the lady at the hearings (Hoermann R et al., Is Pituitary Thyrotropin an Adequate Measure Of Thyroid Hormone-Controlled Homeostasis During Thyroxine Treatment? Eur J Endocrinol 2013;168:271–80).

There was a number of issues raised during the hearing which I found quite interesting to write to you. I believe the case presented has some special aspects but also broader implications. I'd like to briefly comment on thyroid function testing including the TSH-

centred approach that is currently predominant, the role of guidelines and the urgent need for reconsideration of an individualised medicine.

It is a common misconception that the result of a reliable test will also be true.

As an example, let's assume the reliability of a test be 99.99%, which is excellent. With this test, you are screening for a rare condition, which occurs at a low prevalence of 1 in 10,000 in the population. The test result turns out to be pathological in a given patient.

What is the probability that this patient truly suffers from the disease? The question is scientifically answered by applying the Bayes' theorem. A simplified version goes like this, the reliable test fails in 1 in 10,000 cases, and the disease exist in 1 in 10,000 case. That is a ratio of 1 to 1. That means despite its high reliability of 99,99% the test will give a true result in only about 50% under these conditions. You could use a dice to get the same accuracy. The odds do dramatically change when the prevalence of the disease in the population is higher.

What can be done in situations where the test doesn't help to establish a clear diagnosis.

One has to rely on clinical symptoms. If symptoms of the disease are present that obviously will change the probability and the interpretation of the test result.

Any test result has to be interpreted in context and cannot be interpreted without knowing the clinical background and the history of the patient or doing a thorough evaluation of symptoms of the disease (which determines the a priori probability).

A fragmented approach of relying on labs that provide an interpretation of test results without knowledge of the condition of the patient or guidelines that are meant to be generally applicable are in violation not only of common sense, but of fundamental mathematical principles. The availability of a reliable test obviously is a prerequisite, which is met by modern thyroid tests, but does not substitute for the correct interpretation by a physician who knows the patient.

The sad thing is society invest enormous efforts and amounts of money on the development of tests or drugs, but very little in the adequate interpretation of test results and the training of the professionals involved. The solution can only be re-focussing on the individual patient and its unique situation, and a better training in the art of being a doctor.

It is not the guidelines that are critical, but how the guidelines are being applied. Reliability of an assay and validity of a diagnosis are fundamentally different categories. The latter by far transcends the former.

Spending ever more resources on the wrong path like attempting to develop marginally better tests or to write some more precise guidelines does not solve the problem.

Repeated testing without a clue carries a high risk for sick people being misclassified healthy or healthy people being labeled sick (Crowell JM et al., Cumulative Incidence of False-Positive Results in Repeated, Multimodal Cancer Screening. *The Annals of Family Medicine* 2009;7:212–22). Mislabeled not only has a real impact on the life of people, but produces sky-rocketing follow-up costs.

Aren't guidelines a good thing? Yes and No. Guidelines aim at standardisation of an approach or treatment and an equality in care. However, if everybody thinks alike, nobody is left to think different. If nobody thinks different, nobody thinks at all.

Total standardisation is as bad as no standardisation. Preservation of diversity fosters stability at the system level. Strict guidelines and their adherence do not stimulate, but suppress thoughtful and intelligent diagnosis-making. In thyroid function testing, "intelligent diagnosis", as opposed to merely following standard procedure, necessitates using all available tests if and where they are appropriate. Thyroid and adrenal testing are all but two examples that were brought to the attention of Parliament. This is just the tip of an iceberg.

What helps is to train doctors in their people skills and use their talents and intelligence rather than having them merely playing by the rules and robotically adhering to guidelines. We do need a reconsideration of a caring personalised medicine that once existed. This will also require a change of the legal system, which is focused on accountability, rules and regulations to an extent which is no longer healthy and protective, but sick and destructive to any individualism, creativeness and intelligence in society.

The ramifications of this line of thinking go far beyond the thyroid extending to issues like national security or food testing. The blind application of screening tests will not solve these problems either - that is mathematically impossible and will not be changed by well-intentioned, but ill-advised political actions. The road to success is to raise public awareness and to act quickly and intelligently in case of suspicion. The question is not what may be right or wrong in absolute terms that don't exist in the real world, but what is appropriate under the given circumstances.

Back to the thyroid, in my opinion, it is not so much a matter of improving thyroid tests and guidelines that may be required to remedy the situation, but a better understanding and application of the context sensitive nature of the interpretation of those tests that is widely lacking. If society wants to see effective results, it does make a difference where to spent

the money. The ladies have certainly raised an issue that the system may not be working as it should. I have attempted to explain why it may not be all that surprising that it does not work given the way it is designed. The whole approach has to be reconsidered and the resources have to be redirected by equally acknowledging medicine as an art and a technical speciality. Teaching of clinical skills in the diagnosis and treatment of patients and the patient physician relationship are equally important to the implementation of guidelines.

The balance has been lost in recent time, owing to unrealistic expectations by an uneducated public and also the legal demands that are being put on the medical profession. The cost balance is also tilted. No matter what amount of money will be spent, it will never be sufficient, as all attempts to defy the laws of nature and mathematics will eventually fail. Common sense doesn't cost much, but has become a rare commodity. More emphasis needs to be given to improve the adequate training of doctors, and less to the production of ever more useless rules. It is a question of balance, not of either or.

Apart from the general aspects, there are some considerations that specifically relate to the use of TSH in thyroid testing. The currently TSH-centered protocol in thyroid testing has been widely adopted due to its putative ease of use by any medical practitioner, not only specialists in the field of endocrinology, and also their cost effectiveness. The problem with this approach is that TSH is not a statistical parameter like any other laboratory value. First, it is an indirect measure which reflects the supersensitive response of the pituitary gland to circulating thyroid hormones. Secondly, TSH does not follow a normal distribution and displays an asymmetric inverse relation to free thyroxine, which is produced by the thyroid gland. Thirdly, the reference range for TSH, which is obtained in a population, is approx. twice as wide as the normal span in an individual (Andersen S et al., Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. *J Clin Endocrinol Metab* 2002;**87**:1068–72). That means this parameter is not well suited to define and establish true "normality" in a single patient - it is available only as a broader population based reference system.

What this means is that the TSH value that is optimum for an individual patient cannot be reliably and readily obtained by the TSH test- no matter how highly accurate the test is. The true "normal" value can only be derived and interpreted based on the additional consideration of complaints, assessment of symptoms and experienced clinical examination of the patient. The limitations of TSH measurement as a diagnostic and therapeutic target have recently been reviewed by others and myself (Hoermann R &

Midgley JEM. TSH Measurement and Its Implications for Personalised Clinical Decision-Making. *J Thyroid Res* 2012;2012:1–9).

It is important to understand the difference that exists between the individual situation which is not precisely measurable and the statistical average which relates TSH to all sorts of outcomes including mortality. People fail to comprehend - and I must admit it is not that easy to understand - that if there is a report that a low TSH is harmful to health on average, this is not “their TSH” and the same value that is detrimental to one person could save another person’s life. The value doesn’t mean much by itself. There is no one size fits all approach with TSH, as there is with other parameters, which makes it more difficult to deal with.

As a consequence, I would suggest laboratories should not be allowed to provide interpretations of thyroid test results to patients at all, as this just adds to the confusion, since a personal examination is required to correctly interpret a test result and to put it into perspective. GPs with adequate training - which in many countries, however, does not appear to be in place right now - might be able to interpret the thyroid test results they get back from the lab in less complex or uncomplicated cases. For more complex cases, like the ones presented at the hearing, the only solution I can envisage is an adequate referral system from the GP to an endocrinologist and even further to a specialised university department, if needed. The need has to be established on clinical grounds rather than by reliance on laboratory values.

Particularly, in thyroid testing, the objective blood tests and their communication to the patient imply a certainty which does not exist in reality. Although thyroid test are generally very precise and reliable from a technical point of view, I hope I made it clear that the real challenge is with their interpretation, which is highly individual. It is by no means as easy and straightforward, as widely perceived or communicated. Guidelines cannot overcome the high individual variation of TSH that is intrinsic to this parameter and entirely different from other lab values. Standardised guideline cannot provide the extent of flexibility and individuality that is required and cannot substitute for the expertise that is needed in interpreting the test results. The correct interpretation heavily depends on taking into account the patient’s history and subjective complaints as well as the clinical judgment and a physical exam that require specialised knowledge and experience and it sometimes involves trial and error.

While diagnosis remains a challenge, optimisation of thyroid hormone replacement presents even more of a challenge to the patient affected and the doctor who is guiding

the treatment. It is important to recognise apparent limitations of TSH in this area, described e.g. by our group in a recent article (Hoermann R et al., Is Pituitary Thyrotropin an Adequate Measure Of Thyroid Hormone-Controlled Homeostasis During Thyroxine Treatment? *Eur J Endocrinol* 2013;168:271–80). Drug manufactures are also called upon in this area to replace the current levothyroxine drugs that are given as a single bulk dose in the morning by more sophisticated slow-release preparations that are capable of supplying thyroid hormones in a more physiological manner and in amounts of both T3 and T4 that resemble the natural production of the hormones by the human thyroid. It is important to note in this context, that the interchangeability of levothyroxine preparations and many other drugs is not given. That a generic drug contains the identical compound to the original brand doesn't mean it is equally effective. There are other aspects apart from the substance itself like the intestinal resorption that are unique to each preparation and may affect its efficacy. This has been well documented for different thyroid brands and many other drugs (e.g. Carswell JM et al., Generic and Brand-Name L-Thyroxine Are Not Bioequivalent for Children With Severe Congenital Hypothyroidism. *J Clin Endocrinol Metab* Published Online First: 21 December 2012. doi:10.1210/jc.2012-3125).

I think we cannot talk about improving all sorts of things such as diagnostic skills, guidelines, tests and drugs and in the end totally mess up with the implementation of health care policies that may be well intended in their goals including the saving of costs, but are ill-advised and poorly implemented.

I would see supporting clinical quality and advancing human skills as key to both intelligent decision-making and cost effectiveness. The problem is those are soft qualities not measurable quantities. Most quality control programs are focussed on process quality and not aiming at outcome measures. Clearly, what is needed is more practical training, more experience, less regulations, fewer guidelines. The story of the ladies tells it all, as one of them actually suggested that with her personal experience she would be able to diagnose the condition in other people. She didn't say by having read some guidelines.

In case you may have any further questions I could answer I would be happy to do so.

Yours sincerely

Rudolf Hoermann MD