

History of Thyroid Testing

There have been long, sad and unsatisfactory developments in thyroid function testing, including up to the present day.

The first thyroid function test, in the form such tests are used today, appeared in 1960. This measured total thyroxine (T4). Before this, convenient measurement of thyroid hormones was not possible. However, breakthrough though this was, it was immediately realised that this was insufficient for accurate estimation of thyroid function.

Thyroid hormones (T4 and T3) leave the thyroid gland and in the bloodstream are bound onto transport proteins that convey the hormones to the tissues. There are three of these transport proteins: thyroxine-binding globulin (TBG), transthyretin and albumin. Of these TBG is the most important in the average person. It transports about 70% of T4 and 60% of T3.

As the transport proteins and their T4/T3 load pass by the tissues in the bloodstream, very small amounts of hormone are freed as required. These are the free T4 and free T3 fractions. As the tissues remove T4 and T3 for their own use, more is released by the transport proteins for the next tissues to use. The free T4 (FT4) and free T3 (FT3) fractions are a very small percentage of the total circulating hormones. In the case of FT4 in the average person it is about 2/100 of 1% of the total T4 and for FT3 2/10 of 1% of the total T3. Therefore, it is necessary to measure FT4 and FT3 rather than total T4 or total T3.

The problem is that we are all unique in the makeup and amounts of our transport proteins. In the vast majority of people, the TBG levels can be different by at least a factor of 2; and the same (independently) for the other two proteins. There are people with either no TBG at all, or 4 times the normal amount. Their reservoirs of T4 and T3 are therefore hugely different for the same FT4 and FT3. Also, the pregnant woman has twice the TBG and $\frac{3}{4}$ the amount of albumin she had when not pregnant. We also lose transthyretin and albumin when critically ill or with trauma like burns or septicaemia.

To try to get a measure of FT4, a test was developed in 1963-65 to try to convert the total T4 result to a FT4 result. This was the thyroid hormone uptake test. In conjunction with a total T4 result, the two tests could be amalgamated to produce what was claimed was an estimate of FT4. The test method is still used today; e.g. in certain American private labs and elsewhere. However, it is not based on sound principles and does not work properly, especially for people with extreme differences in TBG from the average. Even the pregnant woman's results are compromised.

In the remainder of the 1960's, commercial firms were set up to provide readymade tests for the clinical chemistry labs to use.

In about 1975, commercial TSH and T3 tests were developed and sold. The TSH test was first generation – that is, it could only measure and detect hypothyroidism (the depressed levels in hyperthyroidism were too low to be measured directly). Such was the growing demand for tests that the various companies competed with one another for business in the labs. Since the method of measurement (radioactivity) was the same in all tests, competition was such that no company would have a monopoly of business in the labs. This competition produced faster and slicker tests with shorter and shorter times – giving quicker turnover and more tests done in a given time.

In the late 1970's the shortcomings of the thyroid hormone uptake test, arising from the variation in TBG levels in patients, were very apparent. The demand for properly formulated and soundly developed FT4 and FT3 tests was very great.

As a response, companies and individuals produced various forms of tests claiming to measure these fractions. Many of the offerings were not soundly based, and slowly disappeared into obscurity and obsolescence. Two methods did however prevail and form the basis of FT4 and FT3 testing today.

The London researcher (now a distinguished professor – Nobel Laureate just failed), who had developed the pioneering total test 20 years earlier invented a validly based test for FT4. At the same time, I invented and my company developed and offered a method based on a different principle, but also soundly based.

My method as initially developed was not perfect – there were obscure areas of thyroidology where there were problems but we'd identified them and given advice to circumvent.

The professor's method was sound but suffered from the fact that there were several steps to take before you got an answer, which took time and cost precision – the more handling, the more progressive errors creep in.

On the other hand, the test I had invented was, in the hands of the lab technician exactly the same in handling as the existing total T4 test – a big time/turnover/precision advantage for the busy lab.

The London professor and his group decided to try to destroy the validity and reputation of the rival test and those who had developed it. So began a long series of aggressive, long and detailed theoretical arguments as to why the test I had invented was, in its present form, unfit for purpose and could not and did not work.

In vain did we show that the practical working of our test bore no resemblance whatever to his theoretical predictions – this only invited more and more vituperative denunciation. This aggressive, acrimonious and almost libellous controversy (the worst in the history of any discipline in clinical chemistry) continued for almost 20 years before dying out in the futility with which it had started.

During this time, the average clinical chemistry worker in the average hospital was totally oblivious to all this rarefied argument and was happy that at last a reliable

FT4/FT3 test was available. For example it brought into the diagnostic fold even the most TBG-extreme people mentioned earlier. For a while, there was a golden age in thyroid diagnosis where all tests (TSH, FT4, FT3 were used – especially in Germany and Japan).

In the mid-eighties, pressures on the clinical chemistry lab were beginning to be overwhelming. Such was the demand for tests that the disposal of radioactive waste was too great for licencing of disposal. Consequently non-radioactive detection methods had to be substituted. Two things happened around 1985.

First, second and third generation TSH tests were developed – now one could directly detect both hypo and hyperthyroidism. Secondly, the manufacturers produced several solutions to the nonradioactive detection methods, and integrated them into dedicated automatic analytical platforms. Now one had machines that took the place of the skilled hands-on technician – it was a case now of loading the machine, programming it and pressing the “start” button.

This led to lab monopoly – having chosen the machine, one was confined to the tests dedicated to that machine. However, the individual solutions of the manufacturers to the method of detection in tests led to problems with FT4 and FT3 test development (uniquely).

Unlike all other tests, FT4 and FT3 tests demand special and essential requirements. They must be run at blood temperature (37 degrees), they must sample only a tiny quantity of the available T4 and T3 so as not to sample the T4 and T3 bound to the transport proteins, they must use the same chemical surroundings (for example, salt content, phosphate content) as is present in blood, and they must work in the right acidity as present in blood.

The failure of the development scientists to understand these special requirements, and the compromises needed to make the detection methods work, led to great variation in the performance of the FT4 and especially the FT3 tests between manufacturers offerings. For FT4 this is at present up to 40% difference and for FT3 60%. I would expect no more than a 5% difference as a reasonable variation.

As a result, sensitive TSH tests began to have a paramount position in thyroid function testing. There exists a paradigm of thinking today which closely links FT4 and TSH as a constant relationship over the whole thyroid function spectrum. Therefore, if you do a TSH test, then why do an FT4 test because the TSH value implies an FT4 value – the FT4 test is controversial and inconsistent so why do it? The seeds of TSH only screening had started to sprout.

In 1988 I and my colleagues invented a new test for FT4 and FT3, based on the invention of 1980 but getting rid of the problems at the margins mentioned earlier. Shortly after, I left the field entirely for 10 years, only returning by accident in 1999.

On returning to the field I found it in chaos. In 1992, a group of American scientists had begun to analyse and dissect the commercial FT4 tests to understand why they were so inconsistent. They began a series of papers in the peer reviewed important leading journals which lasted until 2009.

Their findings were on the surface devastating – that is, they alleged that however it came about, all FT4 tests were influenced by the levels of transport proteins in blood – devastating because this meant that they were subject to the T4 and T3 bound by those transport proteins – and the whole point of doing FT4 and FT3 tests is to be independent of these effects.

As it turned out, the whole of this work was completely invalid and wrongly conceived from beginning to end – a completely meaningless study programme. I and a colleague pointed this out but, especially in America, their findings are accepted and further confuse today's understanding of the FT4 and FT3 tests. Meanwhile, the cheap, easy to understand, rapid, and eminently automatable TSH test was gaining strength as a catch-all screen.

In 2005 a new group of US workers came on the scene with a specialised technique for measuring FT4 and FT3 which they alleged was superior to the commercial tests in that it more closely correlated FT4 and TSH.

In 2009 I looked into their work and found it had been done at the wrong temperature – this is important because T4/T3 binding to TBG is very temperature sensitive. On advising them of this, they merely obfuscated and blustered, and though henceforward using the right temperature, did not retract their earlier wrong work but actually included it in papers when they used the right temperature as if the wrong work somehow backed them up – scientific honesty?

Now we come to the present day. We have simultaneously in existence licensed, manufactured and used in diagnosis, tests based on the discredited thyroid hormone uptake test, tests based on sound methodology but including the earlier imperfect tests up to the modern improved ones, and tests offered that are to be run invalidly at room temperature.

This implies a complete failure to regulate by the international regulators whose job it is to ensure equivalence of results. The composite failure of the manufacturers to produce consistent FT4 and FT3 tests has already been mentioned. The failure of the medical thyroidology fraternity to ensure consistency of the tests they use is an additional factor in the diagnostic chaos that is now present. No wonder TSH only screening has gained credence in such an atmosphere.

There is a triple failure that has led to a diagnostic hiatus that urgently needs correcting. The paradigm of the TSH-FT4 relationship is wrong, especially in treatment. The whole conceptual thinking behind diagnosis thyroidology and the importance of personal diagnosis based on the patient rather than whether the numbers fall in or out of the normal range is fatally flawed. For the moment mechanical thinking has traduced medical diagnosis.