

# Should we still be doing duplicate immunoassays?

EVA LESTER,\* CATHRYN CORNS†

From the \*Department of Chemical Pathology, North Middlesex Hospital, London and †University College and Middlesex School of Medicine, London

**SUMMARY** To determine whether, with improvements in radioimmunoassay techniques, duplication is still necessary, the differences between duplicate results for a range of assays done routinely over one month were examined retrospectively. Differences over 10% between duplicates were found in 104/779 (13%) of assays for thyroid stimulating hormone, 27/180 (15%) for total thyroxine, 44/378 (12%) for cortisol, 15/355 (4%) for follicular stimulating hormone, 20/356 (6%) for luteinising hormone, and none for  $\alpha$  fetoprotein (0/256). In only two of 779 patients (0.26%) would the different result of a pair of thyroid stimulating hormone duplicates have led to different courses of action by the laboratory. None of the other differences in any assay would have resulted in a potential misclassification. Although replication of assays will give more correct results by pure scientific criteria, the improvement is rarely clinically important and the financial cost is considerable.

In the past it has been regarded as good scientific practice to carry out clinical chemical assays in duplicate to draw attention to mistakes in the assay. With the automation of most clinical chemistry tests from the introduction of continuous flow analysis in the late 1950's, in most cases duplicate assays have been replaced by single analyses. The basis of a change from duplicate to single analyses has not always been rational; for example, some imprecise manual assays such as amylase are usually only done singly. In most laboratories the only tests which are still carried out in duplicate are radioimmunoassays.

When radioimmunoassays first came into general use in the late 1960's they were difficult tests requiring considerable technical skill, and duplicate or even triplicate assays were mandatory. Twenty years later, with advances in our understanding of the theoretical aspects, improvements in reagents and automation of some or all the stages of the assays, such duplication may no longer be necessary.

This paper looks at the results of duplicate radioimmunoassays carried out in these departments to determine what variation there is between duplicates and whether the use of single assays would cause any clinical problems. The assays examined were for thyroid stimulating hormone (TSH), total thyroxine ( $T_4$ ), follicular stimulating hormone (FSH), luteinising hormone (LH), cortisol and  $\alpha$  fetoprotein. We looked retrospectively at one month's routine work for each assay.

## Methods

### METHODS REVIEWED

Radioimmunoassays in common use can be divided arbitrarily into three categories:

- 1 Entirely manual procedures, with removal of the free fraction by decantation from a solid phase, antibody-bound fraction. Errors are likely to be greatest with this type of assay.
- 2 Semiautomated methods, using automatic pipetting stations, but requiring manual handling at some stage, with a decantation step to separate a liquid free fraction from a solid phase bound fraction.
- 3 Manual, or semiautomated methods using coated beads or tubes, with a minimum potential for errors at the separation stage.

Thyroid stimulating hormone and cortisol assays both use coated tubes or wells and therefore fall into group 3; follicular stimulating hormone, luteinising hormone,  $T_4$  and  $\alpha$  fetoprotein assays all require separation of free fraction from a particulate solid phase.

### LOCAL METHODS

The hormone assays in both these laboratories are carried out by routine medical laboratory scientific officers as part of the general laboratory rotation, under the supervision of a chief or senior medical laboratory scientific officer. Alpha fetoprotein assays are always done by one of two experienced biochemists.

The basis of the methods used is shown in table 1.

Table 1 *Analytical methods*

Analyte	Type of assay	Technique	Supplier	Pipetting system
Thyroid stimulating hormone	IRMA	Coated well	Amersham International	Manual
Total thyroxine	RIA	PEG—assisted double antibody	NETRIA	Micro-Medic II
$\alpha$ fetoprotein	RIA	PEG—assisted single antibody	Department of reproduction physiology, St Bartholomew's Hospital	Manual
Cortisol	RIA	Coated tube	Immunodiagnosics Ltd	Kemtek 1000
Luteinising hormone	RIA	Double antibody, solid phase second antibody	Chelsea Hospital	Kemtek 1000
Follicular stimulating hormone	RIA	Double antibody, solid phase second antibody	Chelsea Hospital	Kemtek 1000

Counting is performed with an NE1600 gamma counter; this is used with a data reduction package which calculates the results for the unknowns. The program also calculates the percentage spread between the duplicates from the formula:

$$\frac{(\text{Higher count rate} - \text{lower count rate})}{\text{Mean count rate}} \times 100$$

This method of calculating the difference between duplicates from count rates rather than on the final concentrations has the disadvantage that even in well optimised assays, lower differences between count rates may give rise to higher differences in the final concentrations, so that the imprecision is not represented adequately.

## Results

Table 2 shows the distribution of the differences between duplicates. In the radioimmunoassay methods the greatest errors occurred at very low and very high analyte concentrations; as a result, even where the precision of the duplicates was worse than 15%, no potential misclassification of patients would have occurred had one of the pair of follicular stimulating hormone, luteinising hormone, or cortisol results been reported, rather than the mean.

For thyroid stimulating hormone, there were five patients in whom one result of a pair of duplicates would have classified them as euthyroid and the other as hyperthyroid. In three cases, however, both results

were below the "action limit" (for this department) of 0.5 mU/l, and a  $T_4$  would have been measured on the basis of either result. In only two cases (0.26%) of the total 779 patients would the different results have led to different courses of action by the laboratory, the results being 0.23 and 0.56 mU/l, and 0.29 and 0.60 mU/l.

No potential misclassification between high (hypothyroid) and normal (euthyroid) thyroid stimulating hormone results occurred.

For  $T_4$ , there were three of 180 patients (1.67%) in whom one result would have classified them as hyperthyroid and the other as euthyroid (171 and 134 nmol/l, 161 and 147 nmol/l, 179 and 157 nmol/l, respectively). All of these patients, however, already had a low thyroid stimulating hormone to alert the requesting clinician.

For  $\alpha$  fetoprotein, there were no patient samples in which the difference between duplicates was greater than 10% and there was no misclassification of patient results. Although our  $\alpha$  fetoprotein assay is manual, so that in theory, the largest errors might be expected, in practice, in the hands of only two skilled operators it is our most precise assay.

## Discussion

It is well established that the worst precision of radioimmunoassay method is at high and at low concentrations and that the design of a good radio-

Table 2 *Analysis of assay errors*

	Thyroid stimulating hormone	Total thyroxine	$\alpha$ fetoprotein	Cortisol	Luteinising hormone	Follicular stimulating hormone
<i>Differences between duplicates:</i>						
Number of batches	19	3	5	7	6	6
Total patient samples	779	180	256	378	356	355
<i>Differences between duplicates:</i>						
< 5%	498 (64)	93 (52)	234 (91)	254 (67)	274 (77)	284 (80)
5-10%	177 (23)	60 (33)	22 (9)	80 (21)	62 (17)	56 (16)
10-15%	37 (5)	13 (7)		28 (7)	12 (3)	11 (3)
> 15%	67 (9)	14 (8)		16 (4)	8 (2)	4 (1)

### Should we still be doing duplicate immunoassays?

immunoassay requires that maximum precision is achieved at clinically important concentrations.<sup>1</sup> The "precision profile" introduced by Ekins,<sup>2</sup> has been widely adopted as a form of quality control in radioimmunoassay as it permits the precision to be calculated at different doses from the duplicate analyses carried out during routine assays. Ekins<sup>2</sup> has also shown that the precision profile is improved by doing duplicate or even triplicate analyses, thereby offering a means of improving the working range of an assay.

Detailed examination of the results obtained in this investigation shows that for luteinising hormone, follicular stimulating hormone, and cortisol the poorest duplication of results was obtained, as predicted, at very high or very low concentrations of hormone. At such concentrations, the relatively poor precision is of little clinical importance and single assays would have provided an equally useful analytical result.

The IRMA assay for thyroid stimulating hormone also has its poorest precision at low hormone concentrations, partly due to limitations of the assay and partly as a result of the greater counting error at low counts. If thyroid stimulating hormone is used as the front line test of thyroid function the clinical validity of the results at the lower end of the reference range can be maintained by performing a second line test, such as total or free T<sub>4</sub> on all samples with a thyroid stimulating hormone below a predefined limit. This limit is currently set at 0.5 mU/l; if single assays are used this should be adjusted to 0.75 or even 1.0 mU/l to ensure that clinically misleading results are not issued by the laboratory. An alternative thyroid function testing strategy would be to measure total T<sub>4</sub> and thyroid stimulating hormone on all patient samples, singly, thereby providing a built-in check on results, while still saving on overall reagent and technical costs.

These results suggest that for many routine radioimmunoassays, methods are adequately precise for singleton assays to be performed, without increasing the likelihood of issuing clinically misleading results, provided that the assays are properly established and controlled. For assays requiring prior extraction of sample, such as testosterone, the need for duplicate assays probably remains.

The case of  $\alpha$  fetoprotein assays is more difficult. As the precision of the  $\alpha$  fetoprotein assay was the best of those examined it would seem that single assays should be recommended. On the other hand,  $\alpha$  fetoprotein is measured as part of a screening programme on all pregnant patients rather than in patients where there are symptoms and signs of endocrine disease. In

addition, the consequences of incorrect  $\alpha$  fetoprotein result are likely to be more disastrous both medically and legally than with endocrine disease where the clinical features or progression of the disease will usually lead to a repeat request. At this stage, therefore, we feel that it is wiser to continue with duplicate  $\alpha$  fetoprotein assays.

It is common practice for standards to be assayed with more replicates than unknowns. This is illogical as standards can be used to fix each other because their relation is known.<sup>1</sup>

All clinical chemists strive for improvements in the work of their departments, but it is necessary to balance costs against benefits to arrive at a realistic standard of performance. The assay of routine thyroid function tests and gonadotrophin samples singly would lead to a saving in reagent and technical costs, and with carefully set "action limits," would lead to a potentially misleading report being issued in less than 0.5% of cases. In one of our laboratories we have changed our thyroid function testing strategy. We now do single thyroid stimulating hormone and total T<sub>4</sub> assays in all patients instead of one duplicated front line test followed by a second test also done in duplicate when the initial test was outside our action limits. For us this has reduced the number of patient tubes assayed by about 25%. As well as the saving in reagent cost, this strategy has the advantage of reducing the amount of technical time in performing assays and more particularly in sorting specimens. With two separate tests, the mathematical probability of misclassifying a patient is reduced which counterbalances the increased probability resulting from single rather than duplicate assays. Dynamic tests such as cortisol and growth hormone responses to stimulation or suppression are also done singly because a single aberrant result in series would almost certainly be detected.

It is very difficult to draw general conclusions. Each laboratory needs to assess its own precision, staff and reagent costs, and its relationship with clinicians for the assays that it performs in duplicate to judge whether they could reasonably be performed singly.

#### References

- 1 Jeffcoate SL. *Efficiency and effectiveness in the endocrine laboratory*. London: Academic Press Inc, 1981:44-71.
- 2 Ekins RP. The "precision profile": its use in RIA assessment and design. *The Ligand Quarterly* 1981;4:33-44.

Requests for reprints to: Dr Eva Lester, Consultant Chemical Pathologist, North Middlesex Hospital, Academic Centre, Sterling Way, Edmonton, London N18 1QX, England.