Chapter 18

The Reliability of Estimating Carbon-14 in Biological Samples

B. Scales

I.C.I. Pharmaceuticals Division, Alderley Park,
Macclesfield, Cheshire, England

INTRODUCTION

Although very many methods now exist for the preparation of samples to be assayed by liquid scintillation counting, most laboratories use relatively few. The reasons for using any one sample preparation technique are often very basic ones, such as cheapness, familiarity, universality and possibly ignorance of other more appropriate methods. On occasions, the reliability of accepted methods is questioned[1] and large discrepancies can be found between results obtained using different methods of sample preparation.

In most laboratories it is common practice to 'normalise' the raw counting data, i.e. to attempt to correct for the unavoidable variables which occur in liquid scintillation counting, by computing d.p.m. values relative to (a) a standard which may be purchased for that purpose or (b) a sample of radioactivity, a portion of which is used in the actual experiment. However, some of the variables cannot be readily assessed, and it is this difficulty which gives rise to the discrepancies which may be found when different methods are used to analyse any one sample.

Further problems can arise when vastly different types of samples have to be analysed, and the results totalled, or compared to give ratios. In these circumstances, partly because of the necessary use of different methods of sample preparation, reproducible, but highly inaccurate final results, can be obtained. The problem can even be magnified and errors increased if different individuals combine their efforts.

The purpose of this paper is to report the results of a study which was done in order to check the agreement between different methods of sample preparation and between different people using the same method. To this end, 22 individuals with different degrees of familiarity with liquid scintillation counting were each asked to assay two different urine and two different plasma samples for their Carbon-14 content, each person using his preferred methods of sample preparation. They were asked to report the results as d.p.m. Carbon-14/ml sample. The only way that the measurements could hopefully differ from normal procedures was that individuals would, where possible, do replicate analyses rather than just one or two. This would enable a more acceptable statistical analysis of the results to be carried out.

METHODS

Preparation of radioactive biological fluids

A high activity urine sample was obtained from a human dosed with ^{14}C-propranolol, and a low activity sample was obtained from a marmoset dosed with ^{14}C-ICI 55, 897. The radioactivity in both these samples was present as stable, highly water-soluble metabolites. The high and low activity human plasma samples contained ^{14}C- ICI 54, 594, a highly protein-bound carboxylic acid.

The approximate specific activities of these samples and the volumes given to each investigator are shown in Table 1. The high activity samples were prepared at about 50, 000 d. p. m. /ml since this is, in general, a typical high level found in urine and plasma samples after a dose of Carbon-14 labelled drug. The low activity samples were of such a specific activity that the background contribution to the count-rate would be significant.

Table 1. Samples prepared for analysis.

Sample	Code	Specific acitivity (approx. d.p.m./ml)	Volume supplied to investigators (ml)
Human urine	H.U.	50, 000	2.0
Marmoset urine	L.U.	500	17.0
Human plasma	H.P.	40, 000	2.5
Human plasma	L.P.	1, 000	12.0

All the specific activities were such that a 10-min count would produce reasonable accuracy, and our standard ^{14}C-hexadecane spiking fluid (nominally 10, 000 d. p. m. /50 µl) would be adequate for efficiency determination after a further 1 to 10-min count.

Sample preparation for liquid scintillation counting

In the event, the Carbon-14 activity in the samples was measured using the three different techniques listed in Table 2.

Table 2. The composition of scintillation media used to assay the samples.

System	Scintillator composition
Emulsion counting:	
Toluene–Triton X-100	Butyl-PBD (0. 8% w/v) in toluene 2 parts + Triton X-100 1 part
Solution counting:	
Dioxane–naphthalene	Butyl-PBD (0.8% w/v) + naphthalene (10.4% w/v) in dioxane
Combustion Techniques:	
Packard 305	CO_2 absorbed in a mixture with final composition, ethanolamine:methanol:butyl-PBD (2. 0% w/v) in toluene (4:9:6).
Kalberer–Rutschmann	Ethanolamine:methanol:butyl-PBD (0. 8% w/v) in toluene (3:22:25)

The emulsion counting system was essentially that described by Turner,[2] and the solution counting system was that described by Werbin et al.[3] The Packard 305 is a semi-automatic flask combustion system which can be operated in either the Hydrogen-3 or the Hydrogen-3 plus Carbon-14 mode, thereby allowing the separation of the two isotopes prior to counting. The second combustion system was

essentially that described by Kalberer and Rutschmann,[4] with minor modifications. The scintillation medium for each of these systems had the composition given in Table 2.

Counting procedures

Samples were counted in either a refrigerated Philips Liquid Scintillation Analyser, refrigerated Packards (models 3340 or 3003) or in room temperature Packards (models 2240). All samples were allowed to reach temperature equilibrium in the counters before the Carbon-14 activity was measured. Samples spiked with ^{14}C-hexadecane (50 μl) were recounted using the same instrument settings.

Categorising results for analysis

The biological samples were counted for from 10 to 100 min and the spiked samples for 1 or 10 min. Although the accuracy of the d.p.m. calculations depends on the net counts accumulated, the variations in calculated d.p.m. caused by the different times used in this study were insignificant. For uniformity, the d.p.m. were calculated from the maximum counts accumulated in any one counting operation. Assuming that the prepared samples were stable, neither the type of instrument used nor the temperature of the counting chamber should affect the d.p.m. calculation. In almost all instances, after examining the prepared samples, these variables could be ignored in categorising the results for analysis.

The most important aspects of the study which affected the d.p.m. calculations were the actual methods of sample preparation. These were categorised for the analysis after due consideration of the physical principles involved and a knowledge of their possible effects.

Thus, all investigators who used the toluene–Triton X-100 system used 10.0 ml of scintillation fluid, but the total aqueous volume ranged from 0.025 to 4.0 ml. The physical properties of this emulsion system were so clearly defined that the prepared samples were readily sub-divided into the categories listed in Table 3.

Table 3. The analysis of urine and plasma samples using the toluene– Triton X-100 system.

Category	Total aqueous volume added to 10 ml scintillation fluid
T-T, 1	Less than 0.25 ml
T-T, 2	From 0.5 to 1.0 ml
T-T, 3	4.0 ml
T-T, 4	4.0 ml of L.P.
T-T, 5	as T-T, 2, but no b.g. subtract for L.P. and L.U.

A number of variations in the use of the dioxane– naphthalene (Table 4) system were considered for categorisation. However, the results obtained using 12.5% aqueous volume in a total of from 14 to 20 ml at refrigerated or ambient temperatures were statistically identical to those obtained using 5% aqueous volume at ambient temperature. The addition of fumed silica (Cabosil, Degussa Aerosil) was also without effect. All these methods are listed as D-N, 1. The only misuse of the method was in the failure by one person to add sufficient water to prevent crystallisation in the 0-4° C machines. This method was therefore considered separately (D-N, 2). The range of sample volumes used in the two combustion systems is also given in Table 4. Both systems were relatively free from obvious investigator error. Only one investigator used the Kalberer-Rutschmann method, and because of the obvious differences in the absorption systems, his results were considered separately (C, 2).

Table 4. The analysis of urine and plasma samples using the dioxane-naphthalene system and the combustion techniques.

Category	% Aqueous volume and conditions		
D–N, 1	12.5% at 0°C – room temp.	)	± fumed silica
	5% at room temp.	)	
D–N, 2	5% at 0–4°C		
C, 1	Packard 305	)	0.02 to 0.50 ml
C, 2	Kalberer–Rutschmann	)	on low ash paper

RESULTS AND DISCUSSION

Statistical analysis of the results

All the individual figures cannot be shown. A summary of results for each of the nine categories of sample preparation is given in Table 5.

The first line of figures in each compartment (a) gives the number of investigators, the number of batches and the total number of samples.

The second line (b) gives the calculated mean specific activity (d.p.m. Carbon-14/ml) and an asterisk indicates that it is significantly different ($P < 0.05$) from the overall mean at the foot of the table.

The third line (c) gives the 'within person' coefficient of variation (C.V. %). This is a measure of the repeatability of the method on that sample when used by one person. It is the standard deviation of the observations expressed as a percentage of the mean, where the standard deviation is based only on the variation encountered when one person makes repeat determinations.

The fourth line (d) gives the 'between person' C.V. %. This is a measure of the reproducibility with which different people using the same method on one sample can obtain similar results. It is calculated in the same way as the previous figure, except that the standard deviation is based on the variations encountered when different individuals make repeat determinations. An asterisk after this figure indicates that the value is significantly greater ($P < 0.05$) than the previous one; this shows that there are significant differences between people using the same method.

A question mark in each compartment indicates that the values were obtained from relatively little data and should be interpreted with care.

A visual examination of the prepared samples in the counting vials, and of the raw counting data obtained from them, showed that certain results listed in Table 5 could be sensibly ignored in the calculation of the overall results. Thus the results from method T - T, 4 on L.P. and D - N, 2 on H.U. could be discarded because of gross inhomogeneity of the samples, resulting in very unstable and irreproducible counting rates. Also, the results from T - T, 5 could be eliminated because of failure to subtract background counts.

The statistical effect of discarding these obviously unreliable results was a minor one, the result being to slightly raise the overall mean specific activity and to slightly lower both of the C.V. %'s. The high overall 'between person' C.V. % was therefore caused by less obvious errors than these.

Individual and method differences

The mean results obtained by each individual were tested to see if they were statistically different from the overall mean for that sample. Of the 22 investigators in the study, 10 produced results on all 4 samples which were indistinguishable from the overall means. These 10 investigators used mainly methods T - T, 1,

T - T, 2 and D - N, 1.

Table 5. Mean specific activities and coefficients of variation for each sample and each method.

Method		Sample			
		L.P.	H.P.	L.U.	H.U.
T–T, 1	(a)		4–5–24		6–6–24
	(b)	–	39,085*	–	50,652
	(c)		2.0		1.9
	(d)		3.5*		2.3
T–T, 2	(a)	9–15–71	6–8–41	9–11–50	7–8–44
	(b)	1405*	39,218*	546	52,536*
	(c)	5.2	1.7	3.6	2.1
	(d)	13.3*	2.6*	5.9*	4.1*
T–T, 3	(a)	3–3–6	4–6–15	4–4–9	1–1–2
	(b)	1462*	30,924*	440*	39,800*
	(c)	? 3.7	3.6	10.5	? 4.2
	(d)	3.9	10.4	16.7	–
T–T, 4	(a)	1–1–3			
	(b)	832*	–	–	–
	(c)	3.0			
	(d)	–			
T–T, 5	(a)	1–1–2		1–2–8	
	(b)	1481*		629*	
	(c)	0.9	–	? 3.1	–
	(d)	–		28.3*	
D–N, 1	(a)	5–8–20	6–7–19	7–8–22	7–8–28
	(b)	1532	37,697	550	51,131
	(c)	1.4	3.0	3.4	4.5
	(d)	4.8*	13.9*	18.3*	8.5*
D–N, 2	(a)				1–1–4
	(b)				? 42,875*
	(c)	–	–	–	12.2
	(d)				–
C, 1	(a)	5–5–17	5–6–14	3–3–7	3–3–6
	(b)	1230*	34,386*	413*	46,034*
	(c)	6.0	12.1	? 6.9	? 1.7
	(d)	16.5*	14.3	14.6	11.2*
C, 2	(a)	1–1–3			
	(b)	1422*			
	(c)	? 0.4	–	–	–
	(d)	–			
Overall	(b)	1367	37,235	535	50,868
	(c)	4.9	4.3	4.6	3.6
	(d)	12.2*	10.2*	14.4*	6.6*

(a) = No. of investigators — batches — samples.
(b) = Mean specific activity (d.p.m. Carbon-14/ml).
(c) = Within person C.V.%.
(d) = Between person C.V.%.
? = Few data available, interpret with care.
* on line (b) means that the result is significantly different ($P<0.05$) from overall values.
* on line (d) means that this value is significantly greater than result in line (c).

However, 8 other investigators each produced 3 or 4 results which were significantly different from the overall means ($P < 0.05$), and they used mainly methods T - T, 3, D - N, 1, C, 1 and to a lesser extent T - T, 2. As can be seen from Table 5, these methods produced the highest 'between person' C.V. %, although it is important to note that method D - N, 1 was the only method to give mean specific activities for all 4 samples which were indistinguishable from the overall means. After careful examination of the raw data and discussion with the participants it was decided that in most, but not all, cases it was the analytical systems employed, rather than the manipulative techniques of the individuals, which were at fault.

Examination of raw data from emulsion counting

Method T - T, 1. This method of sample preparation was only used with the high activity samples, Table 5. It was differentiated from T - T, 2 because it had been observed in preliminary studies that the use of small volumes of water could result in a separation of droplets containing water-soluble salts which assayed with a different efficiency to solvent-soluble materials. It was hoped that the use of highly water-soluble ^{14}C-Inderal metabolites and ^{14}C-54, 594 would amplify this point. In the event, the mean result for H. P. was significantly above the overall mean, and that for H. U. was indistinguishable from the mean. Only with the high activity urine sample did different people obtain essentially the same result.

Method T - T, 2. This was the most commonly used method. Only with the low activity urine sample did the mean specific activity agree with the overall mean (Table 5); the results from the other three samples were all significantly higher than the overall mean.

On a number of occasions, the samples in a batch showed marked differences in optical properties, either before or after spiking. This was very noticeable with diluted plasma samples, and counting efficiencies were found to range from 33 to 42%. These changes in physical properties of the samples were ignored, and there resulted a wide range of calculated specific activities.

In contrast, the use of 1.0 ml undiluted low activity plasma occasionally gave rise to an unexpectedly high and constant counting efficiency of from 61 to 62%. This resulted in a relatively low specific activity. These observations were disturbing because they showed that presumably inaccurate results could be obtained with good repeatability.

Method T - T, 3. This method was used on all samples. In the majority of cases the low activity samples were diluted at least four-fold and the high activity samples forty-fold, as part of the procedure of adding 4.0 ml aqueous phase to 10.00 ml scintillator.

With the low activity urine samples, three participants used 4.0 ml undiluted urine. The efficiency of one of these batches of samples was only 9% due to inadequate channel settings, but the result was acceptable.

The mean results by this method for H.P., L.U., and H.U. (Table 5) are 18 – 24% below the overall average values. An examination of the raw data showed that the counting efficiencies were quite reproducible and that it was variability in the initial 10-min count rate which was responsible for the large 'within person' coefficient of variation, i.e. the lack of repeatability.

In general, in all categories of emulsion counting, the high 'between person' C.V.% was due primarily to instability of samples. This resulted in unstable sample counts and variable counting efficiencies, which occurred in at least 30% of the prepared batches. In not one instance did an investigator assay different dilutions of sample to see if he would obtain the same result.

Examination of raw data from solution counting

Method D - N, 1. This was the second most common method used. The mean results for each of the four samples were indistinguishable from the overall means (Table 5), but as with most other assay methods the 'between person' C.V.% was relatively high. The 'within person' C.V.% was similar to that of method T - T, 2, and further examination of individuals' raw data was undertaken to try to explain the differences between people.

The method was used by some of the least experienced personnel who, in retrospect, used incorrect pipetting techniques. Also they spiked only one sample in each batch, even though the count rate of the samples varied by up to 8%. Moreover, one investigator failed to completely dissolve the hexadecane spike in the dioxane scintillator, thereby obtaining a false value for the counting efficiency.

These problems apart, the method gave very good reproducibility. Only one person assayed a range of volumes for each sample. The average specific activities were all very close to the overall means and the standard deviations of the results were as low as any. This illustrates that in this system, even with varying sample volumes, the repeatability is high.

Therefore, the large 'between person' C.V.% observed in solution counting was caused primarily by incorrect sample preparation due to poor training.

Examination of raw data from combustion counting

Method C, 1. This was the third most popular method, and as shown in Table 5 it was used for all samples. Not only are the mean values for each of the four samples significantly lower than the overall means (24% for L.U.), but the 'within person' and the 'between person' coefficients of variation are excessively high for almost all samples.

Although combustion to CO_2 and H_2O is hailed by many as the only reliable method for biological samples, many combustion systems leave much to be desired. The repeated blocking of the gas lines leading from the combustion vessel, with carbonaceous material, is ample proof that many combustions are incomplete.

On some occasions, very high blank values were noted when non-labelled samples were combusted immediately after a labelled sample; in these cases, carry over was equivalent to 5% of the Carbon-14 count of the previously combusted sample.

Although the machine did occasionally meter a variable volume of liquid into different vials, this should not seriously affect the d.p.m. calculations when, as was the case in this study, all samples were spiked.

Therefore, the gross variability and the overall low values were due to:

(a) the inadequate combustion of many samples, and
(b) the incomplete transfer and collection of combustion products.

Method C, 2. Only one individual used this method, and obtained the highest specific activity and the lowest 'within person' C.V.% for all the combusted samples of L.P. Such a result, in contrast to method C, 1, is probably due to complete dissolution of combustion products in the scintillator, whether they were fully oxidised to CO_2 or not.

SUMMARY

For each sample, it has been shown that there are highly significant differences between methods ($P < 0.005$) and between people using the same method ($P < 0.005$). The particular differences vary from sample to sample.

Methods which rely on the solubility of the aqueous sample in dioxane gave results for each of the four samples which were indistinguishable from the overall

means. The majority of the results obtained with toluene– Triton X-100 after the addition of no more than 10% aqueous solution were significantly higher than the overall means, whereas when the water content was greater than 10% the majority of the results were significantly lower.

Variations in solution counting were chiefly due to inaccurate pipetting, leading to large 'between person' coefficients of variation, whereas variations in most of the emulsion systems were due largely to sample instability under counting conditions.

The combustion of samples by the Packard 305 automatic system gave very unreliable results in that, besides giving results well below the overall means, the machine produced the largest 'within person' and 'between person' coefficients of variation of any of the methods tried. The Kalberer – Rutschmann variation appeared to be the most reliable of the two combustion procedures at present available.

On the whole, these results are very disturbing, and are all the more difficult to assess because no absolute standard exists. It is probable that the most reliable and acceptable method is that which gives the smallest 'within person' and 'between person' coefficients of variation.

Table 6. Reliability of methods.

Method	No. of batches assayed	% of batches giving mean d.p.m. statistically indistinguishable from overall means	Reliability of method
T–T, 1	11	100	good
T–T, 2	42	95	good
D–N, 1	31	77	good
T–T, 5	3	66	bad
T–T, 3	14	50	bad
C, 1	17	30	bad
C, 2	1	0	good
T–T, 4	1	0	bad
D–N, 2	1	0	bad

As shown in Table 6, and when the above considerations are taken into account, methods T - T, 1, T - T, 2 and D - N, 1 are to be preferred. Although we always set ourselves the highest standards for accuracy and reliability in our experiments, it seems that, as carried out above, all leave something to be desired.

ACKNOWLEDGEMENTS

I would like to thank all those who contributed to this study. I am also grateful to Mr. J.A. Lewis for the statistical analysis of the results.

REFERENCES

1. W.A. Cope, and J.A. Double, in Liquid Scintillation Counting, Vol. 2, (Ed. M.A. Crook, P. Johnson and B. Scales), Heyden and Son, London, 1972, p.223.
2. J.C. Turner, Intern. J. Applied Radiation and Isotopes **19**, 557 (1968).
3. W. Werbin, I.L. Chaikoff and M.R. Imada, Proc. Soc. Exp. Biol. and Med. **102**, 8 (1959).
4. F. Kalberer and J. Rutschmann, Helv. Chim. Acta **44**, 1956 (1961).

DISCUSSION

F.R.Jacobsberg: I presume scintillation counting is still carried out at ICI Pharmaceuticals. What are your own conclusions regarding techniques?

B. Scales: Yes, we continue to do a great deal of liquid scintillation counting. In my talk I have painted the very worst picture, but this extreme case can exist. Our problem is possibly due to the fact that we do not offer a radioisotope analytical service, but let everyone loose on our many machines, each person taking full responsibility for his own results. One possible outcome of this is shown in the above results and we are attempting to alter this situation. My own beliefs are that when assaying any sample, replicates should be assayed, either by using different volumes of sample or by using different scintillation systems. In these circumstances, the agreement of replicate specific activities is indicative of a reasonably reliable analysis.