The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

1939-16

Joint ICTP-IAEA Workshop on Nuclear Structure and Decay Data: Theory and Evaluation

28 April - 9 May, 2008

Background Information for Data Analyses (Convergence of techniques for the evaluation of discrepant data)

Desmond MacMahon

National Physical Laboratory
London
U.K.

3B2v7.51c GML4.3.1 ARI : 2739 Prod.Type:COM pp.1–7(col.fig.:1–3) ED:Mariamma PAGN: nvs SCAN: raghu

ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Applied Radiation and Isotopes ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Applied Radiation and Isotopes

www.elsevier.com/locate/apradiso

Convergence of techniques for the evaluation of discrepant data

Desmond MacMahon*, Andy Pearce, Peter Harris

National Physical Laboratory, Queens Road, Teddington, Middlesex, TW11 0LW, UK

Abstract

The problem of evaluating discrepant data has been addressed by several authors over the previous 20 yr. More recently some attention has been given to the use of the median, which is expected to have better statistical 'robustness'. The various evaluation techniques should converge towards the 'true' value as the number of data in a data set increases, and the 'robustness' of each evaluation technique can then be tested by the rate at which that technique converges. Several evaluation techniques have been applied to discrepant data sets, and the results are shown to converge as the size of the data set grows. The discrepant data sets used as examples are the measured half-lives of ^{90}Sr and ^{137}Cs. Differences in the behaviour of the evaluation techniques are discussed, as applied to these data sets. The half-lives deduced from this study are: ^{90}Sr 10551 $\pm$ 14 days; ^{137}Cs 10981 $\pm$ 11 days.

Keywords: Data evaluation; Discrepant data

1. Introduction

A significant problem faced by any data evaluator is to determine the best method of deriving a recommended value and an associated uncertainty from a discrepant set of data. This difficulty has been addressed by several authors with particular reference to radionuclide half-life data, and a number of data evaluation procedures have been proposed in recent years (Zijp, 1985; Woods and Munster, 1988; Gray et al., 1990; Woods, 1990; James et al., 1991; Rajput and MacMahon, 1992; Kafala et al., 1994; Müller, 2000; Helene and Vanin, 2002; Cox, 2002).

The statistical techniques developed for the evaluation of discrepant data sets may be summarised as follows:

1.1. *Limitation of relative statistical weights (LRSW)*

Zijp (1985) proposed that no single datum should have a relative statistical weight greater than 0.50 when determining the weighted mean of a data set. The uncertainty of any datum which did should be increased until its relative statistical weight is reduced to 0.50. Woods and Munster (1988) further proposed that the unweighted mean of the data set and the new weighted mean should be compared. If their uncertainties overlapped, the weighted mean should be adopted. If their uncertainties did not overlap, the data were inconsistent and it would be safer to use the unweighted mean. In either case the uncertainty quoted would be inflated, if necessary, to include the value of the data set with the lowest uncertainty.

1.2. *Normalised Residuals*

James et al. (1992) introduced an evaluation technique in which the uncertainties of only discrepant data were adjusted. Such discrepant data are identified on the basis of their normalised residuals (R_i), defined as

$$R_i = \sqrt{\frac{w_i W}{(W - w_i)}}(x_i - \bar{x}),$$

where the weighted mean $\bar{x} = \sum x_i w_i / W$, $\quad w_i = 1/\sigma_i^2$

*Corresponding author.

E-mail address: desmond.macmahon@npl.co.uk (D. MacMahon).

0969-8043/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.apradiso.2003.11.028

and $W = \sum w_i$, x_i and σ_i are the measured values and their associated uncertainties, respectively.

A limiting value of the normalised residual (R_0) for a set of n values is defined as

$$R_0 = \sqrt{1.8 \ln N + 2.6} \quad \text{for } 2 \leqslant N \leqslant 100.$$

If any value in the data set has $|R_i| > R_0$, the weight of the value with the largest R_i is reduced until the normalised residual is reduced to R_0. This procedure is repeated until no normalised residual is greater than R_0. The weighted mean is then recalculated with the adjusted weights.

1.3. Rajeval

As proposed by Rajput and MacMahon (1992), this technique shares the same basic principle as that of James et al. (1991) in that the uncertainties of only the more discrepant data are adjusted. The technique comprises of three stages:

(i) Outliers in the data set are detected by calculating the quantity y_i

$$y_i = \frac{x_i - x_{ui}}{\sqrt{\sigma_i^2 + \sigma_{ui}^2}},$$

x_{ui} is the unweighted mean of all the data set excluding x_i, and σ_{ui} is the standard deviation associated with x_{ui}. The critical value of $|y_i|$ is 1.96 at 5% significance level for a two-tailed test. Measurements with $|y_i| > 3 \times 1.96$ are considered to be outliers and may be excluded from further stages in the evaluation;

(ii) Inconsistent measurements that remain in the data set after the population test are revealed by calculating a standardised deviate Z_i:

$$Z_i = \frac{x_i - \bar{x}}{\sqrt{\sigma_i^2 - \sigma_w^2}},$$

where

$$\sigma_w = \sqrt{\frac{1}{W}}$$

for each Z_i the probability integral

$$P(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) \mathrm{d}t,$$

is determined. The absolute difference between $P(z)$ and 0.5 is a measure of the central deviation (CD). A critical value of the central deviation (cv) can be determined by the following expression:

$$\mathrm{cv} = [(0.5)^{N/(N-1)}] \quad \text{for } N > 1;$$

(iii) If the central deviation of any value is greater than the critical value, that value is regarded as inconsistent. The uncertainties of the inconsistent values are adjusted to σ_i':

$$\sigma_i' = \sqrt{\sigma_i^2 + \sigma_w^2}.$$

An iteration procedure is adopted in which σ_w is recalculated each time and added in quadrature to the uncertainties of those values with CD $>$ cv. The iteration process is terminated when all CD $<$ cv.

1.4. Median

The median of a set of data is rather insensitive to outliers and has recently been regarded as a more robust method of evaluating a discrepant data set. The question arises as to what uncertainty to associate with the median. Müller (2000) has suggested that use is made of the median of the absolute deviations (MAD), where

$$\mathrm{MAD} = \mathrm{med}\{|x_i - \tilde{m}|\} \quad \text{and} \quad \tilde{m} = \mathrm{med}\{x_i\}.$$

The uncertainty of $\tilde{m}$ is then taken as $s(\tilde{m}) = (1.858 \times \mathrm{MAD})/\sqrt{n}$.

The median is a robust estimator but, as it takes no account of the uncertainties associated with the individual values in the data set, some of the information content of the input data is lost.

1.5. Bootstrap Method

Helene and Vanin (2002) have proposed a Bootstrap Method, based on a Monte Carlo procedure, to estimate a best value and associated uncertainty. A random sample (with replacement) is selected and the median $x_{\mathrm{med},j}$ is determined from a set of experimental data $\{x_i\}$ $(i = 1, 2, \ldots, n)$. After repeating the sampling for $j = 1, 2, \ldots, M$, the best estimate of the quantity is given by

$$\hat{x} = \frac{1}{M} \sum_{j=1}^{M} x_{\mathrm{med},j}$$

with variance

$$\sigma_{\hat{x}}^2 = \frac{1}{M-1} \sum_{j=1}^{M} (x_{\mathrm{med},j} - \hat{x})^2.$$

Note that each sample, j, may have some values of the data set repeated and other values missing. As in the case of the simple median, the Bootstrap Method does not make use of the uncertainties quoted with the data.

1.6. Extension to the Bootstrap Method

Cox (2002) has described a procedure based on the median, but also making use of the quoted uncertainties. If the only information available is the measured half-life and associated standard uncertainty, a Gaussian distribution is assigned to that input quantity. Random samples are then taken from the probability distribution for each of the input quantities. About one million Monte Carlo trials are recommended. The recom-

Table 1
Cs-137 Half-life data

Authors	Measured half-lives		Weighted mean		LRSW		Normalized residuals		Rajeval		Median		Bootstrap		Extended bootstrap	
	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ
Wiles and Tomlinson (1955a)	9715	146	9715	146	9715	146	9715	146	9715	146	9715	146	9715	146	9715	146
Brown et al. (1955)	10957	146	10336	103	10336	621	10336	103	10336	103	10336	816	10336	439	10336	103
Farrar et al. (1961)	11103	146	10592	84	10592	877	10993	102	11045	113	10957	160	10673	570	10928	127
Fleishman et al. (1962a,b)	10994	256	10631	80	10631	916	10989	94	11025	96	10975.5	68	10798	367	10932	121
Gorbics et al. (1963)	10840	18	10830	18	10736	220	10845	27	10904	68	10957	97	10873	296	10898	87
Rider et al. (1963)	10665	110	10826	17	10741	161	10840	28	10841	18	10898.5	125	10840	203	10850	68
Lewis et al. (1963)	11220	47	10873	16	10930	120	10891	93	11031	74	10957	103	10914	170	10901	86
Flynn et al. (1965a)	10921	183	10873	16	10928	109	10892	82	11006	68	10939	86	10919	115	10905	74
Flynn et al. (1965b)	11286	256	10875	16	10931	102	10909	80	11041	62	10957	90	10958	107	10940	89
Harbottle (1970)	11191	157	10878	16	10936	96	10944	77	11073	55	10975.5	67	10992	100	10978	81
Emery et al. (1972)	11023	37	10901	15	10934	94	11011	45	11030	30	10994	88	10998	82	10992	67
Dietz and Pachucki (1973)	11020.8	4.1	11012	4	10961	60	11020	7	11021	4	11007	87	11002	62	11002	45
Corbett (1973)	11034	29	11013	4	10975	46	11021	7	11021	4	11021	53	11008	51	11013	33
Gries and Steyn (1978)	10906	33	11011	4	10973	48	11020	7	11021	4	11007	64	10996	47	10998	34
Houtermans et al. (1980)	11009	11	11011	4	10996	25	11018	6	11019	4	11009	46	11000	40	11005	25
Martin and Taylor (1990)	10967.8	4.5	10994	3	10994	27	10987	13	10996	10	11001.5	42	10995	33	10998	21
Gostely (1992)	10940.8	6.9	10986	3	10986	35	10969	8	10969	4	10994	41	10988	32	10990	23
Unterweger (2002)	11018.3	9.5	10988	3	10988	32	10988	11	11007	7	10981	25	10994	28	10997	18
Schrader (2004)	10970	20	10988	3	10988	33	10985	10	10970	4	10970	23	10990	26	10992	19

All half-life data and standard deviations are in days.

mended value and standard deviation are then calculated as shown for the Bootstrap Method above.

2. Measurements and evaluations

Table 1 lists all the published values, with uncertainties, of the half-life of ^{137}Cs in the chronological order of their publication. Also shown are the results of applying each of the above data evaluation techniques as each new data point is added to the set. All half-life values and uncertainties in Table 1 are in units of days. The reduced chi-squared for the complete data set of 19 values is 18.6, indicating the existence of significant discrepancies.

Fig. 1 shows the data of Table 1 in graphical form. Fig. 2 shows the latter 9 points of Fig. 1, expanded to show the behaviour of the measured data and the evaluations as they converge.

The same information for the smaller half-life data set of ^{90}Sr is shown in Table 2 and Fig. 3. In the case of this data set, the reduced chi-squared is 40.0.

The intention of this work is to demonstrate how the various methods of evaluating discrepant data converge as the number of points in the data set increases. This is clearly shown in Figs. 1 and 2 for the half-life data of ^{137}Cs. The earliest point is clearly discrepant but it has been retained in the data set to show how the different techniques deal with this problem. From the left-hand side of Fig. 1 it can be seen that the weighted mean, the LRSW and the Bootstrap Methods are strongly

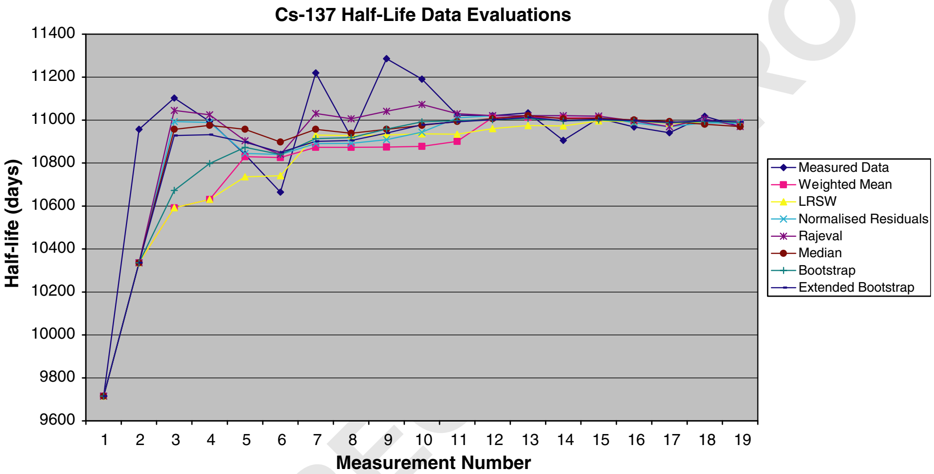

Fig. 1. Cs-137 Half-life data evaluations.

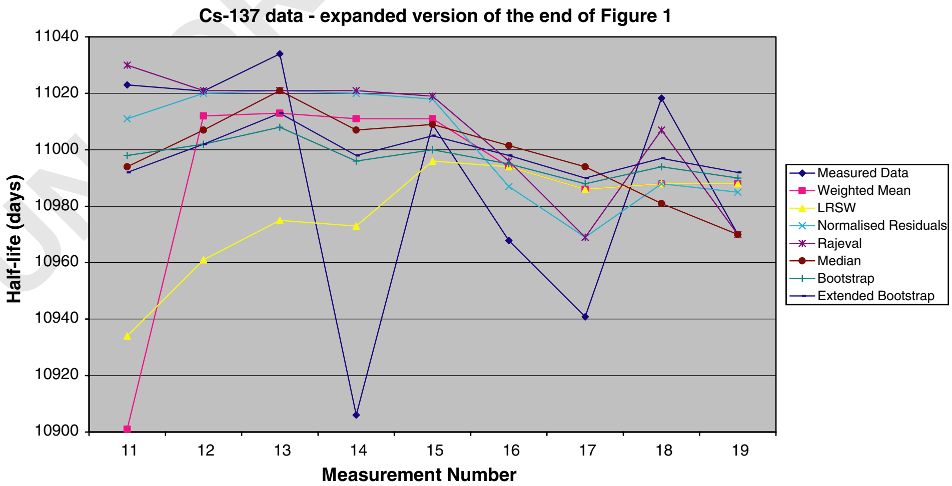

Fig. 2. Cs-137 Data—expanded version of the end of Fig. 1.

Table 2
Sr-90 Half-life data

Authors	Measured half-lives		Weighted mean		LRSW		Normalized residuals		Rajeval		Median		Bootstrap		Extended bootstrap	
	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ	$t_{1/2}$	σ
Wiles and Tomlinson (1955b)	10120	150	10120	150	10120	150	10120	150	10120	150	10120	150	10120	150	10120	150
Anikina et al. (1958a,b)	10700	580	10156	145	10410	410	10156	145	10156	145	10410	381	10410	205	10409	300
Flynn et al. (1965)	10230	150	10192	104	10192	104	10192	104	10192	104	10230	118	10323	227	10235	136
Flynn et al. (1965)	10410	330	10212	99	10212	99	10212	99	10212	99	10320	135	10348	147	10296	150
Hoppes (1977)	10636	88	10450	65	10424	212	10347	84	10404	131	10410	188	10422	177	10379	167
Lagoutine et al. (1978)	10282	12	10287	12	10366	84	10283	12	10282	12	10346	92	10381	133	10338	98
Ramthun (1983)	10588	91	10292	12	10390	108	10314	48	10337	65	10410	126	10426	145	10391	117
Kochin et al. (1989)	10665	37	10326	11	10446	164	10525	69	10573	52	10499	139	10478	132	10459	99
Martin et al. (1994)	10561	14	10418	9	10426	144	10565	23	10563	13	10561	93	10503	123	10507	87
Woods and Lucas (1996)	10495	4	10482	4	10456	39	10542	21	10496	4	10528	81	10504	94	10505	51
Schrader (2004)	10557	11	10489	3	10483	30	10550	14	10552	10	10561	62	10521	82	10528	32

All half-life data and standard deviation are in days.

influenced by the earliest discrepant point until there are at least 6 further measurements. On the other hand, the Normalised Residuals, Rajeval and median reach a value close to 11,000 days after only the third measurement (to avoid congestion in the figures, the uncertainties in the data have not been included; and the first 3 points all have the same uncertainty). Fig. 2 shows how the evaluations converge as the final 9 data points are added to the data set.

The smaller data set for ^{90}Sr is rather different, as shown in Fig. 3. Some convergence of the evaluation techniques is evident only after the last two data points are included. There is a large scatter in the experimental data and there is a worrying general upward trend in the results of the evaluation methods. This trend is clearly evident when using the weighted mean, where a straight line fit to the weighted mean data would indicate that the half-life of ^{90}Sr is increasing by 34 days each time this important parameter is measured! However, with the inclusion of the final data point, there is a spread of only 0.7% in the evaluations.

3. Conclusions

3.1. ^{137}Cs

The ^{137}Cs data displayed in Fig. 1 exhibit the type of behaviour one might have expected, i.e. as measurement techniques improve the scatter in the measured values decreases and the results of the evaluation techniques tend to converge. The left-hand side of Fig. 1 shows that there are significant differences in the ways the evaluation techniques behave with small numbers of discrepant data, with the Median, Normalised Residuals and Rajeval techniques recovering from the influence of the first discrepant point much more quickly than the other techniques. The right-hand side of Fig. 2 shows that, when all 19 points have been included, the Median and Rajeval techniques have converged on a value of 10970 days, while the other techniques have converged on a value close to the weighted mean—10988 days. However, the results of all the evaluation techniques, shown on the bottom line of Table 1, cover a range of only 0.2%. A value of 10981 $\pm$11 days covers the results of all the evaluation techniques and can be adopted as the current best estimate of the half-life of ^{137}Cs.

3.2. ^{90}Sr

The situation with the ^{90}Sr half-life data is much less satisfactory, firstly because the data are more discrepant and secondly because there is a general upward trend in the data. One can only speculate that earlier data may have been affected by undetected shorter half-life contaminants. The curves in Fig. 3 are converging only

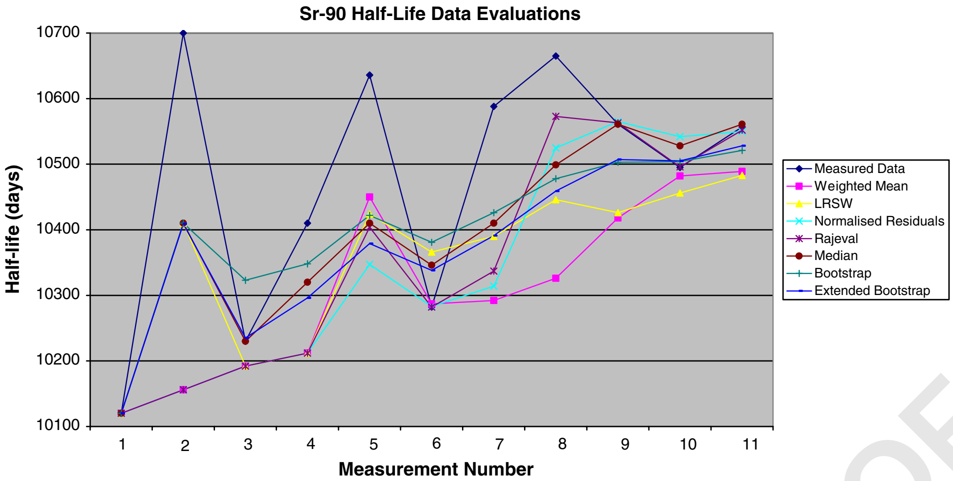

Fig. 3. Sr-90 Half-life data evaluations.

slowly. To have confidence in the evaluated half-life of ^{90}Sr it is essential that further careful measurements are carried out. Pending the results of further measurements, it can be seen from the bottom line of Table 2 that the results of the Median, Bootstrap, Extended Bootstrap, Normalised Residuals and Rajeval techniques are consistent and that a value of 10551 ± 14 days (the mean of the two latter techniques with the larger of the two uncertainties) can be deduced for the current best estimate of the half-life of ^{90}Sr. The weighted mean and LRSW values are significantly lower but are heavily influenced by discrepant values.

References

Anikina, M.P., Ivanov, R.N., Kukavadze, G.M., Ershler, B.V., 1958a. Half-life of Sr90 and its yield in U233 fission. At. Energy USSR 4, 198.

Anikina, M.P., Ivanov, R.N., Kukavadze, G.M., Ershler, B.V., 1958b. Sov. J. At. Energy, 4, 270–271.

Brown, F., Hall, G.R., Walter, A.J., 1955. The half-life of ^{137}Cs. J. Inorg. Nucl. Chem. 1, 241–247.

Corbett, J.A., 1973. Radiometric measurement of the half-life of caesium-137. Nucl. Eng. Int. 18, 715.

Cox, M.G., 2002. The evaluation of key comparison data. Metrologia 39, 589–595.

Dietz, L.A., Pachucki, C.F., 1973. ^{137}Cs and ^{134}Cs half-lives determined by mass spectrometry. J. Inorg. Nucl. Chem. 35, 1769–1776.

Emery, J.F., Reynolds, S.A., Wyatt, E.I., Gleason, G.I., 1972. Half-lives of radionuclides—IV. Nucl. Sci. Eng. 48, 319.

Farrar, H., Dasgupta, A.K., Tomlinson, R.H., 1961. Half-life of ^{137}Cs. Can. J. Chem. 39, 681.

Fleishman, D.G., Burovina, I.V., Nesterov, V.P., 1962a. Half-life of ^{137}Cs. At. Energy USSR 13, 592.

Fleishman, D.G., Burovina, I.V., Nesterov, V.P., 1962b. Half-life of ^{137}Cs. Sov. J. At. Energy 13, 1224.

Flynn, K.F., Glendenin, L.E., Harkness, A.L., Steinberg, E.P., 1965. Half-lives of ^{90}Sr and ^{137}Cs. J. Inorg. Nucl. Chem. 27, 21–23.

Gray, P.W., MacMahon, T.D., Rajput, M.U., 1990. Objective data evaluation procedures. Nucl. Instrum. Methods A 286, 569–575.

Gries, W.H., Steyn, J., 1978. Determination of the half-life of ^{137}Cs. Nucl. Instrum. Methods 152, 459–462.

Gorbics, S.G., Kunz, W.E., Nash, A.E., 1963. New values for half-lives of ^{137}Cs and ^{60}Co nuclides. Nucleonics 21, 63.

Gostely, J.J., 1992. A determination of the half-life of ^{137}Cs. Appl. Radiat. Isot. 43, 949–951.

Harbottle, G., 1970. The half-lives of two long-lived isomers, ^{108m}Ag and ^{192m2}Ir, and of ^{137}Cs and ^{204}Tl. Radiochim. Acta 13, 132.

Helene, O., Vanin, V.R., 2002. Analysis of discrepant data using a bootstrap procedure. Nucl. Instrum. Methods A 481, 626–631.

Hoppes, D.D., 1977, NBS private communication.

Houtermans, H., Milosevic, O., Reichel, F., 1980. Half-lives of 35 radionuclide. Int. J. Appl. Radiat. Isot. 31, 153–154.

James, M.F., Mills, R.W., Weaver, D.R., 1992. The use of the normalised residual in averaging experimental data and in treating outliers. AEA Technology Report AEA-RS-1082.

Kafala, S.I., MacMahon, T.D., Gray, P.W., 1994. Testing of data evaluation methods. Nucl. Instrum. Methods A 339, 151–157.

Kochin, A.E., Kuzmina, M.G., Sokolova, I.A., Merson, P.L., 1989. Measurement of the ^{90}Sr half-life. Metrologia 26, 203–204.

Lagoutine, F., Legrand, J., Bac, C., 1978. Périodes de quelques radionucléides. Int. J. Appl. Radiat. Isot. 29, 269–272.

Lewis, R.E., McHenry, R.E., Butler, T.A., 1963. Half-life of ^{137}Cs. Trans. Am. Nucl. Soc. 8, 79.

Martin, R.H., Taylor, J.G.V., 1990. A measurement of the half-life of ^{137}Cs. Nucl. Instrum. Methods 286, 507–513.

Martin, R.H., Burns, K.I.W., Taylor, J.G.V., 1994. A measurement of the half-life of ^{90}Sr. Nucl. Instrum. Methods 339, 158–163.

Müller, J.W., 2000. Possible advantages of a robust evaluation of comparisons. J. Res. Natl. Inst. Stand. Technol. 105, 551-555 and 781.

Rajput, M.U., MacMahon, T.D., 1992. Techniques for evaluating discrepant data. Nucl. Instrum. Methods A 312, 289–295.

Ramthun, H., 1983. Calorimetric redetermination of the half-life of ^{90}Sr. Nucl. Instrum. Methods 207, 445–448.

Rider, B.F., Peterson, J.P., Ruiz, C.P., 1963. The mass spectrometric determination of the ^{137}Cs half-life. Nucl. Sci. Eng. 15, 284–287.

Schrader, H., 2004. Half-life measurements with ionisation chambers—a study of systematic effects and results. Appl. Radiat. Isot., accepted for publication in this issue.

Unterweger, M.P., 2002. Half-life measurements at the National Institute of Standards and Technology. Appl. Radiat. Isot. 56, 125–130.

Wiles, D.M., Tomlinson, R.H., 1955a. Half-life of ^{137}Cs. Phys. Rev. 99, 188.

Wiles, D.M., Tomlinson, R.H., 1955b. Half-life of strontium90. Can. J. Phys. 33, 133–137.

Woods, M.J., Lucas, S.E.M., 1996. Half-life of ^{90}Sr—measurement and critical review. Nucl. Instrum. Methods A 69, 534–538.

Woods, M.J., Munster, A.S., 1988. Evaluation of half-life data. NPL Report RS(EXT)95.

Zijp, W.L., 1985. On the statistical evaluation of inconsistent measurement results illustrated on the example of the ^{90}Sr half-life. ECN-179, Petten, The Netherlands.