On the determination of the B lifetime by combining the results of different experiments

Louis Lyons

Nuclear Physics Laboratory, Keble Road, Oxford, England

Alex J. Martin

H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, England

David H. Saxon

Rutherford Appleton Laboratory, Chilton, Didcot, Oxon, England

(Received 26 June 1989)

The current estimate of the B lifetime is obtained by averaging the results of six experiments. We point out problems connected with the conventional approach to combining data and discuss ways of overcoming them. Our recommended value is 1.13 ± 0.15 ps. Our discussion is relevant to a wide range of problems in which data are combined.

The naive method of combining results $(\tau_i \pm \sigma_i)$ of different experiments is based on minimizing the sum of squares:

$$S(\tau) = \sum \left[\frac{\tau_i - \tau}{\sigma_i} \right]^2 .$$ (1)

It yields a weighted average

$$\hat{\tau} = \frac{\sum \tau_i / \sigma_i^2}{\sum 1/\sigma_i^2}$$ (2)

with error

$$\hat{\sigma} = \left[\sum 1/\sigma_i^2 \right]^{-1/2} .$$ (3)

This is, for example, the technique used by the Particle Data Group.[1] We point out that this averaging procedure is not unique and probably not optimal. Although most of our comments are of much more general applicability, we illustrate our discussion by specific reference to the case of the B lifetime.

There are several problems which complicate the use of the above procedure.

(i) The errors (σ_i) are supposed to be the *true* errors on each experiment, whereas what are available are the *estimated* errors, which may depend on the lifetime estimate (τ_i). For example, when the experimental resolutions are small, the maximum-likelihood method yields σ_i that are proportional to τ_i. Thus for a series of experiments measuring N_i events, the results $(\tau_i \pm \tau_i/\sqrt{N_i})$ together with Eq. (2) yield an estimate

$$\hat{\tau} = \frac{\sum N_i / \tau_i}{\sum N_i / \tau_i^2} .$$ (4)

This violates what we call "the combination principle": the combining technique should give the same answer as if all the data had been regarded as a single experiment,[2,3] viz.,

$$\hat{\tau} = \frac{\sum N_i \tau_i}{\sum N_i} .$$ (5)

(ii) Lifetime errors are often asymmetric, whereas Eqs. (2) and (3) use a single number for the error. This problem is acute for experiments with low statistics and is hence relatively unimportant for the B-lifetime results we wish to combine.

(iii) Experiments have systematic as well as random errors and often the systematic errors are correlated between experiments. Thus uncertainties in assumptions about the fragmentation of b quarks, or in the lifetime of possible charm contamination in the sample, would produce common effects on several experiments. These difficulties can be compounded by the use of contradictory assumptions by different experiments.

(iv) Experiments use different techniques for extracting lifetimes (e.g., maximum likelihood or trimmed means;[3] decay distances or impact parameters).

(v) Resolution effects on the measured decay distance or impact parameter may very well need to be treated differently from other statistical errors.

(vi) The samples selected by the various experiments may contain different mixtures of B hadrons. There is, however, experimental evidence[4,5] and theoretical prejudice that the different B-hadron species have similar lifetimes. We therefore do not consider this further.

An alternative approach for combining data is to add the log-likelihood functions of the separate experiments. This automatically takes care of asymmetric errors; resolution effects can be incorporated in the individual likelihood functions and the problem of estimated errors is completely absent. The drawback is that there is no ready way to incorporate systematic errors, whether or not they are correlated. It is *incorrect* to use the likelihood functions to obtain the best value and its statistical error, and to combine the systematic errors afterwards. It would be valuable to have a procedure for widening the likelihood function of each experiment in order to allow

for its systematic errors. Another complication is that not all experiments determine lifetimes by likelihood methods. Even for those that have done so, it is often the case that only the result and the upper and lower errors are quoted, from which it is impossible to reconstruct the likelihood unambiguously.

We are thus persuaded to look for some method which uses just the quoted results, but is better than Eq. (2). For statistical errors, we propose that this equation can be used, but with the errors (σ_i) set at the values they would have been had each experimental result been $\hat{\tau}$ rather than τ_i (compare page 26 of Ref. 2, and Refs. 6 and 7). Thus

$$\hat{\tau} = \frac{\sum \tau_i / \sigma_i^2(\hat{\tau})}{\sum 1 / \sigma_i^2(\hat{\tau})} \tag{6}$$

and

$$\hat{\sigma} = \left[\sum 1/\sigma_i^2(\hat{\tau}) \right]^{-1/2} . \tag{7}$$

Equation (6) can be solved iteratively for $\hat{\tau}$. We call our method the "locally matched solution" (LMS). It is similar to the iteratively reweighted least-squares method.[6]

The LMS method is *not* equivalent to minimizing $S(\tau)$ of Eq. (1), with the σ_i regarded as functions of τ. This is because the functional dependence of the error on τ_i is a description of how, for example, the second derivative of the log-likelihood function at its maximum would have varied, had τ_i been different, rather than giving the shape of the log-likelihood function as we move away from the maximum. A consequence is that one cannot interpret $S(\hat{\tau})$ as a χ^2 distribution in order to assess whether the individual lifetime measurements are consistent.

To see the way in which the error estimates scale with τ_i, and to study the properties of the LMS method, we have generated a series of Monte Carlo "experiments," each consisting of 100 decays, generated according to an exponential distribution of lifetime $\tau_0 = 1.0$, smeared with Gaussian resolution functions of width r: i.e.,

$$\frac{dn}{dt} = \frac{1}{\sqrt{2\pi} r \tau_0} \int_0^\infty \exp\left[-\frac{t'}{\tau_0} \right] \exp\left[-\frac{(t-t')^2}{2r^2} \right] dt'$$

$$= \frac{1}{2\tau_0} \exp\left[\frac{r^2}{2\tau_0^2} - \frac{t}{\tau_0} \right] \text{erfc}\left[\frac{1}{\sqrt{2}} \left[\frac{r}{\tau_0} - \frac{t}{r} \right] \right] . \tag{8}$$

We show some of these distributions in Fig. 1.

For each "experiment," we determine the lifetime by a maximum-likelihood fit, and its error by half the difference in the values of τ at which the log-likelihood is $\frac{1}{2}$ less than its maximum. For each value of r, we want to see how the error varies with τ_i. We thus select experiments producing a given value of τ_i, and then find the average of the errors for these experiments. The dependence of this average error on τ_i is shown in Fig. 2; the error bars are estimated from the spread of lifetime errors for experiments in that bin.

From Fig. 2 we are led to a parametrization:

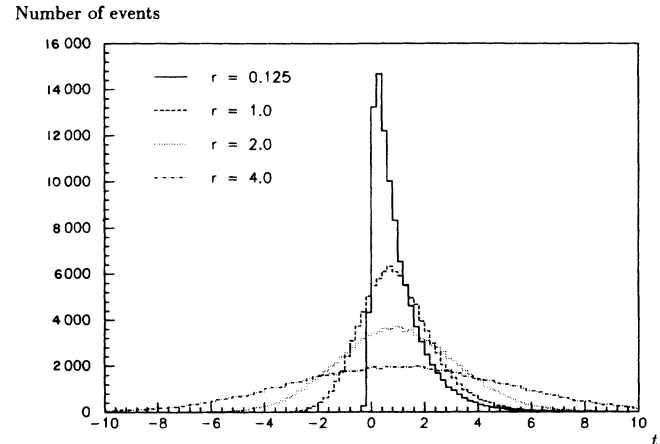$$\sigma_i^2(\hat{\tau}) = \frac{k_i \hat{\tau}^2}{N_i} + \frac{r_i^2}{N_i} , \tag{9}$$

FIG. 1. Monte Carlo decay-time distributions, with lifetime $\tau_0 = 1.0$ and with resolutions $r = 0.125$, 1.0, 2.0, and 4.0. The distributions change from being almost exponential when r/τ_0 is small to being approximately Gaussian, centered on τ_0, and with width r, when r/τ_0 is large.

where k_i is a factor which describes the statistical accuracy of the experimental technique employed (i.e., neglecting resolution), and is unity for direct measurement of the decay times. This is what would have been expected had the τ_i been determined instead from the average of the decay times.[8] For impact-parameter methods, $k_i > 1$, because of the intrinsic smearing produced by the range of decay angles and kinematics. Its value depends on the Lorentz-boost factor of the decaying hadron in the laboratory, and on the momentum cuts. We estimate by a simple Monte Carlo program that $k_i = 1.3$ at the SLAC storage ring PEP (29 GeV) and $k_i = 1.4$ at the DESY storage ring PETRA (35 GeV), and is only weakly sensitive to the momentum cuts. For two-jet methods, looking at the separation of decay vertices, $k_i = 0.5$.

We then apply the LMS method [i.e., Eq. (6), with the parametrization (9)] to pairs of Monte Carlo experiments.
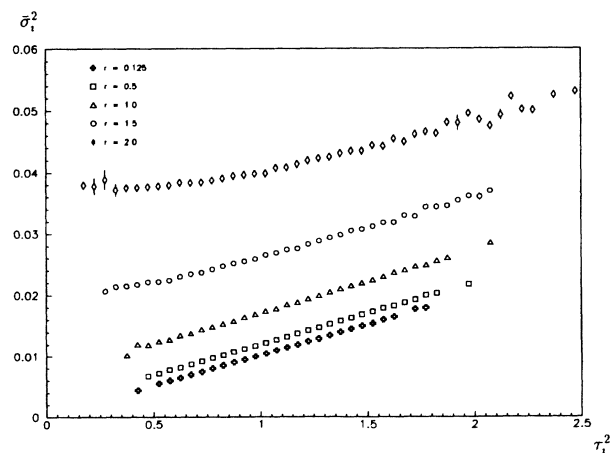

FIG. 2. The dependence of the average estimated statistical error $\bar{\sigma}_i$ squared on the square of the lifetime estimate τ_i, for Monte Carlo "experiments" of 100 events with lifetime $\tau_0 = 1.0$ and with resolutions $r = 0.125$, 0.5, 1.0, 1.5, and 2.0.

TABLE I. Results of the comparison of the LMS method (II) with the naive least-squares procedure (I). These results are from 1000 Monte Carlo trials (or 10 000 for $r=2$) of combining two "experiments," each consisting of 100 events. The Monte Carlo generator uses a lifetime of 1.00, and four separate values of the experimental resolution r (fixed values of 0.5, 1.0, and 2.0, and a uniform distribution in the range 0.5–1.5). The table lists the means of the $\hat{\tau}$ distributions ($\langle\hat{\tau}\rangle$); the means of Δ, where $\Delta=\tau_{\text{fit}}-\hat{\tau}$, and τ_{fit} is the result from a maximum-likelihood fit to a combined sample of 200 events; and the rms deviations from zero of the Δ distributions.

r		$\langle\hat{\tau}\rangle$	$\langle\Delta\rangle$	$\langle\Delta^2\rangle^{1/2}$
0.50	(I)	0.991±0.002	0.010	0.014
	(II)	1.000±0.002	0.001	0.001
1.00	(I)	0.989±0.003	0.010	0.014
	(II)	0.997±0.003	0.002	0.003
0.5–1.5	(I)	0.991±0.003	0.011	0.015
	(II)	1.000±0.003	0.002	0.003
2.00	(I)	0.990±0.002	0.011	0.015
	(II)	0.999±0.002	0.003	0.004

In Table I we compare its performance with that of Eq. (2). We see that it does not suffer from the downward bias of the naive method. We have also investigated the two procedures as far as the combination principle is concerned, by finding how the combined $\hat{\tau}$ for a pair of "experiments" each of 100 events compares with τ_{fit}, the value obtained for a single likelihood fit to all 200 events together. Since for the combination principle these should agree, we define $\Delta=\tau_{\text{fit}}-\hat{\tau}$, and expect $\langle\Delta\rangle$ and $\langle\Delta^2\rangle^{1/2}$ to be zero; we see that again the LMS method is significantly better (see also Ref. 7).

We have also tested the method with the more realistic situation where an experiment has a spread of r values. The case where r is uniformly distributed between 0.5 and 1.5 yields a graph of $\bar{\sigma}_i^2$ against τ_i^2 which is virtually indistinguishable from that for the fixed value $r=1$ of Fig. 2. The satisfactory performance of the LMS method for the variable r case is shown in Table I.

We now turn to the systematic errors and their correla-

tions. If the errors are treated as constants, the combined lifetime is obtained by minimizing with respect to τ:

$$S(\tau)=\sum_{ij}(\tau_i-\tau)(E^{-1})_{ij}(\tau_j-\tau)\ ,\qquad(10)$$

where E_{ij} is the error matrix.[9] This yields

$$\hat{\tau}=\sum_{ij}\tau_i(E^{-1})_{ij}\Big/\sum_{ij}(E^{-1})_{ij}\qquad(11)$$

with error

$$\hat{\sigma}=\left[\sum_{ij}(E^{-1})_{ij}\right]^{-1/2}\ .\qquad(12)$$

This approach has been used by Muller,[10] following an earlier simplified treatment by Wu.[11] When the elements of the error matrix are estimated, we again adopt the procedure of using $E_{ij}(\hat{\tau})$ in the right-hand sides of the above equations.

In order to apply Eqs. (11) and (12) to the B lifetime, we need to specify the functional dependence of E_{ij} on τ_{ij}. We make the following assumptions.

(i) The statistical errors are uncorrelated.

(ii) Statistical errors can be parametrized as in Eq. (9), with the values of k_i and r_i as given in Table II, and the effective N_i chosen to reproduce the quoted statistical errors. Here we have estimated the event measuring errors from the published information on resolutions, mostly from impact parameters, and the scale factor from mean impact parameter ($\langle\delta\rangle$) to mean lifetime ($\langle\delta\rangle\sim0.6c\tau$ at $\sqrt{s}=29-35$ GeV). In the case of TASSO, this error and the scale factor are the result of some averaging over methods.

(iii) The systematic errors are also expressible as $(\alpha_i^2\hat{\tau}^2+\beta_i^2)^{1/2}$ (see Table II). Apart from MAC, the subdivision of systematic errors into scale factors and offsets is our own, based on published information.

(iv) The correlations between all pairs of systematic errors are the same, with correlation coefficient c, which we allow to vary from zero to unity (in principle c could be negative, but we believe it is not likely to be so). (We could have been more careful about the systematic error matrix by considering how each source of error contrib-

TABLE II. B-lifetime measurements. In the "Result" column, the first errors are statistical and the second are systematic. For each experiment, the statistical errors of an individual event are taken as $(k_i\hat{\tau}^2+r_i^2)^{1/2}$ [compare with Eq. (9)]; the systematic ones on the lifetime are parametrized as $(\alpha_i^2\hat{\tau}^2+\beta_i^2)^{1/2}$.

Experiment	Result (ps)	Statistical errors Scale (k_i)	Statistical errors Resolution (r_i) (ps)	Systematic errors Scale (α_i)	Systematic errors Offset (β_i) (ps)
DELCO (Ref. 12)	$1.17^{+0.27+0.17}_{-0.22-0.16}$	1.3	1.6	0.09	0.14
HRS (Ref. 13)	$1.02^{+0.41}_{-0.37}$	1.3	1.7		
JADE (δ_e) (Ref. 14)	$1.27^{+0.35}_{-0.29}\pm0.24$	1.4	1.8	0.17	0.10
JADE (δ_μ) (Ref. 14)	$1.36^{+0.32}_{-0.27}\pm0.21$	1.4	1.8	0.15	0.07
JADE (dipole) (Ref. 15)	$1.46^{+0.22}_{-0.21}\pm0.34$	0.5	1.7	0.20	0.18
MAC (Ref. 16)	(1.29±0.20±0.07) ×(1.00±0.15)	1.3	1.7	0.15	0.07
Mark II (Ref. 17)	0.98±0.12±0.13	1.3	0.9	0.10	0.07
TASSO (Ref. 5)	1.35±0.10±0.24	1.0	1.1	0.16	0.13

utes to the separate experiments. This, however, is a major undertaking, beyond the scope of this study.)

Then for any assumed value of τ, we construct E_{ij} by adding the diagonal statistical and full systematic error matrices, and use it in Eqs. (11) and (12).

We have calculated the combined lifetime $\hat{\tau}$ for the experiments listed in Table II as a function of the common correlation coefficient c for the systematic errors (solid line in Fig. 3). We have compared this with extreme assumptions about the $\hat{\tau}$ dependence of the errors: the naive method using constant errors (dashed line), and the situation where both statistical and systematic errors are proportional to $\hat{\tau}$ (dotted line).

In the absence of detailed knowledge of c, we take its value as 0.5 ± 0.3. This then provides a lifetime estimate of $1.13\pm0.14\pm0.05$ ps, where the first error is statistical and the second arises from the uncertainty in c. We ascribe a further error of ±0.03 ps from the uncertainties of the coefficients of the constant and $\hat{\tau}$-dependent contributions to the error. This gives us our final estimate of 1.13 ± 0.15 ps. Compared with the naive method, the effect of the variation of the errors with $\hat{\tau}$ is to increase $\hat{\tau}_B$ by 0.03 ps, whilst the correlation of systematic errors lowers it by 0.07 ps; the statistical error is increased by 40%. Our answer is dominated by the Mark II result, with its small errors (for example, if we increase their errors by 50%, we obtain $1.21\pm0.16\pm0.03\pm0.03$ ps); the fact that $\hat{\tau}_B$ decreases as c increases is caused by their particularly small systematic error. Our result is close to that obtained by Muller using slightly different input from Table II. By including the effects of correlations he obtained 1.15 ± 0.14 ps (Ref. 10).

Finally, we note that the type of considerations discussed here can be applied to a wide variety of problems of combining data from experiments where the estimated errors depend on the measured values of the parameters. Thus in measuring cross sections for rare processes or small branching ratios, the result is likely to be proportional to $N\pm\sqrt{N}$. A low estimate of N will yield a small apparent error, and again it is necessary to avoid giving such a measurement an unduly large weight.
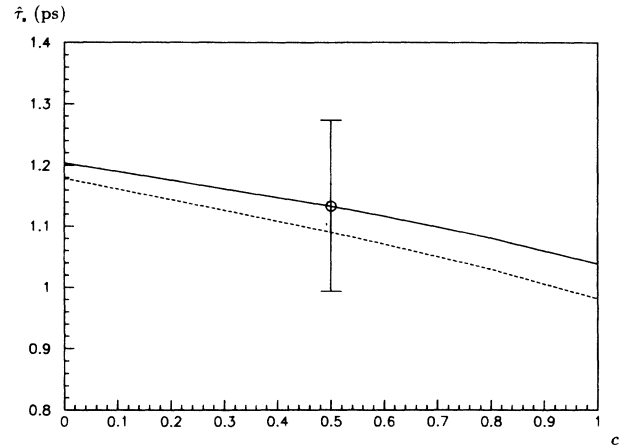

FIG. 3. The dependence of the weighted averaged B lifetime ($\hat{\tau}_B$) on the correlation coefficient c of the systematic errors. The dashed line is for the naive least-squares method with the random and systematic errors constant; the dotted one has both the errors proportional to $\hat{\tau}$, and in the solid one they are parametrized as given in Table II. The point with error bars show our preferred value.

[1]Particle Data Group, G. P. Yost et al., Phys. Lett. B **204**, 1 (1988).

[2]L. Lyons, Statistics for Nuclear and Particle Physicists (Cambridge University Press, Cambridge, England, 1986).

[3]L. Lyons and D. H. Saxon, Rep. Prog. Phys. **52**, 1015 (1989).

[4]CLEO Collaboration, A. Bean et al., Phys. Rev. Lett. **58**, 183 (1987).

[5]TASSO Collaboration, W. Braunschweig et al., Z. Phys. C **44**, 1 (1989).

[6]P. J. Green, J. R. Stat. Soc., B **46**, 149 (1984).

[7]G. P. Yost, Particle Data Group Internal Note No. 88-02 (unpublished).

[8]G. P. Yost, Nucl. Instrum. Methods **224**, 489 (1984).

[9]L. Lyons, D. Gibaut, and P. Clifford, Nucl. Instrum. Methods **270**, 110 (1988).

[10]D. Muller, in Proceedings of the XXIV International Conference on High Energy Physics, Munich, West Germany, 1988, edited by R. Kotthaus and J. H. Kühn (Springer, Berlin, 1989), p. 884.

[11]S. L. Wu, in Lepton and Photon Interactions, proceedings of the International Symposium on Lepton and Photon Interactions at High Energies, Hamburg, West Germany, 1987, edited by R. Rückl and W. Bartel [Nucl. Phys. B (Proc. Suppl.) **3**, 39 (1988)].

[12]DELCO Collaboration, D. E. Klem, et al., Phys. Rev. D **37**, 41 (1988).

[13]HRS Collaboration, J-M. Brom et al., Phys. Lett. B **195**, 301 (1987).

[14]JADE Collaboration, J. Hagemann (private communication).

[15]R. W. H Ramcke, thesis, University of Hamburg, 1988.

[16]MAC Collaboration, W. W. Ash et al., Phys. Rev. Lett. **58**, 640 (1987).

[17]Mark II Collaboration, R. A. Ong et al., Phys. Rev. Lett. **62**, 1236 (1989).