Acta Cryst. (1970). A**26**, 162

Necessity for weighting in Hamilton's *R*-factor ratio test.

By G.C. FORD* and J.S. ROLLETT, *University Computing Laboratory, 19 Parks Road, Oxford, OX1 3PL, England*

(*Received* 23 *December* 1968 *and in revised form* 21 *May* 1969)

It is shown that there are theoretical grounds for believing that ratios of unweighted *R*-factors will be more variable than ratios of properly weighted least-squares minimization functions. An experimental confirmation of this is reported. It is recommended that apparently significantly large unweighted *R*-factor ratios should be checked with proper weighting.

Hamilton (1965) provides tables of *R*-factor ratios, corresponding to different significance levels, for a comparison between two refinements of a crystal structure when one refinement model is related to the other by the addition of extra parameters. The last paragraph of his § 3 suggests that conventional *R*-factors may be used in place of his weighted *R*-factor and in example 4 of § 5 he distinguishes between a conventional and a weighted *R*-factor. This may have led some workers to apply the test to unweighted *R*-factors although Hamilton does not actually recommend this. The temptation to do so is raised by the unfortunately common practice of publishing the final *R*-factor for a crystal structure refinement, but not the minimization function. Comparison of the refinement with others can then only be based on the conventional *R*-factor unless the minimization function is recomputed.

We show that an unweighted *R*-factor test applied to the results of refinements with proper weights can be grossly misleading. (If the refinements themselves are unweighted the accuracy of the whole analysis is suspect unless it can be shown that all the $|F_{o\mathbf{h}}|$ are of equal uncertainty.) The effects are surprisingly large, and we have good reason to believe that crystallographers have not in general realized that this is so.

Let n values of $|F_{o\mathbf{h}}|$ be fitted by means of $|F_{c\mathbf{h}}|$ depending on s parameters. In the notation of Weatherburn (1962) the quantity $\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}(|F_{c\mathbf{h}}|-|F_{c\mathbf{h}}|)^2$ is twice a $\gamma[\frac{1}{2}(n-s)]$ variate with mean $(n-s)$ and variance $2(n-s)$, provided that the weights $w_{\mathbf{h}}$ are properly chosen and on an absolute scale. For a single term $w_{\mathbf{h}}(|F_{o\mathbf{h}}|-|F_{c\mathbf{h}}|)^2$ the expectation value is $(n-s)/n$ and the variance $2(n-s)/n$. Hence we have:

Quantity	Mean	Variance
$Q_1=\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}(\|F_{o\mathbf{h}}\|-\|F_{c\mathbf{h}}\|)^2$	$(n-s)$	$2(n-s)$
$Q_2=\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} (\|F_{o\mathbf{h}}\|-\|F_{c\mathbf{h}}\|)^2$	$[(n-s)/n]\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}^{-1}$	$2[(n-s)/n]\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}^{-2}$

If we write ϱ for the ratio of the estimated coefficient of variation of Q_2 divided by that of Q_1 (the coefficient of variation is the estimated standard deviation divided by the mean) we have

$$\varrho=\sqrt{(n\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}^{-2})/\sum_{\mathbf{h}=\mathbf{h}_1}^{\mathbf{h}_n} w_{\mathbf{h}}^{-1}}\,.$$

This ratio is always greater than unity unless all of the $w_{\mathbf{h}}$ are equal. If for example there are 1000 **h** with $w_{\mathbf{h}}=1{\cdot}0$ and 100 **h** with $w_{\mathbf{h}}=0{\cdot}1$ then $\varrho=1{\cdot}74$, and this is similar to the situation which can arise when the original weighting scheme of Hughes (1941) is appropriate. The ratio of unweighted *R*-factors is likely to be nearly twice as variable as the corresponding ratio of weighted *R*-factors, and it is not obvious that use of $\Sigma||F_{o\mathbf{h}}|-|F_{c\mathbf{h}}||$ rather than $[\Sigma(|F_{o\mathbf{h}}|-|F_{c\mathbf{h}}|)^2]^{1/2}$ will alter this.

An example has confirmed this conclusion. O. Kennard and K.A. Kerr, of the Cambridge University Chemical Laboratory, have kindly provided us with their unpublished data for the structure factors of 1,2,3,4-tetrachloro-5,6-diphenylcalicene (Kennard *et al.*, 1967) and we have carried out refinements of the structure parameters with and without constraints corresponding to various theoretical models for the bond lengths. Waser's (1963) method of augmenting the normal equations was used to apply the constraints which were given a high weight so that deviations from them were negligible in terms of the e.s.d. of the freely refined parameters. Thus each independent constraint reduced by one the number of independent structure parameters. In one case the weighted *R*-factor ratio was 1·005 for a hypothesis with 9 degrees of freedom and an unconstrained refinement with 1578 degrees of freedom. This corresponded to a probability between 0·1 and 0·05, in good agreement with a probability of 0·2 from a χ^2 test on bond-length differences. The constrained model could not be rejected with any significance. The corresponding unweighted *R*-factor ratio was 1·010 which rejected the constrained model at the 0·005 significance level, despite the use of $\Sigma||F_{o\mathbf{h}}|-|F_{c\mathbf{h}}||$ rather than the $\Sigma(|F_{o\mathbf{h}}|-|F_{c\mathbf{h}}|)^2$ of Q_2.

We conclude that, if a Hamilton test with unweighted *R*-factors indicates that the higher *R*-factor differs significantly from the lower, it is necessary to calculate the ratio of the weighted *R*-factors to confirm the conclusion.

We wish to thank a referee for remarks which have improved the clarity of this note.

References

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502.
HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
KENNARD, O., WATSON, D. G., FAWCETT, J. K., KERR, K. A. & ROMERS, C. (1967). *Tetrahedron Letters*, **39**, 3885.
WASER, J. (1963). *Acta Cryst.* **16**, 1091.
WEATHERBURN, C. E. (1962). *A First Course in Mathematical Statistics*. Cambridge University Press.

* Present address: Department of Biological Sciences, Purdue University, Lafayette, Indiana 47907, U.S.A.