17.7-1 A STATISTICAL ANALYSIS OF STRUCTURAL PRECISION FROM CRYSTALLOGRAPHIC DATA. By Frank H. Allen and Michael J. Doyle, Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge, England.

The use of the Cambridge Structural Database (CSD) for systematic analyses of related structures is now a common research technique. Usually only the more precise structures are selected for this purpose. Current versions of CSD record only the R-factor and an AS flag as indicators of precision : AS records the mean esd of C-C bonds ($\bar{\sigma}$) as AS=1 for $\bar{\sigma}$ up to 0.005A, 2 for 0.0051-0.010, 3 for 0.0101-0.030, 4 for 0.0301+. Unfortunately a significant number of papers (16%) do not report relevant esd's and AS=0; furthermore the range initially chosen (in 1970) for AS=3 is somewhat too broad. Observation shows that many simple organics have $\bar{\sigma}$ in the range 0.0101-0.015A and would be acceptable in many projects, whilst values of $\bar{\sigma}$ = 0.02-0.03A may be less desirable.

The R-AS relationship for 33,448 entries from CSD (no disorder, error-free coordinates available) is summarized in the Table. We have also examined the variation of R and AS with : no. of non-H atoms, presence of H as a %age of scattering matter, maximum atomic number (Zmax), and an r.m.s. atomic number normalized to carbon (Z') where $Z' = (\Sigma Z^2/n.Zc^2)^{\frac{1}{2}}$, and n=no. of non-H atoms and Zc=6. The work shows that AS may be expressed as a function of R and either Zmax or Z' for predictive purposes designed to infill AS = 0 entries.

Since 1985, however, individual coordinate esd's have been input to CSD. Values of $\bar{\sigma}$ (in A) have been calculated for 4317 entries and a plot of $\bar{\sigma}$ vs R for five ranges of Zmax is illustrated below. Regressions of the form : $\bar{\sigma} = a + bR + c(Zmax)$ and $\bar{\sigma} = a + b.R.Z'$ have been calculated. The latter equation yields predicted values ($\bar{\sigma}'$) which differ from $\bar{\sigma}$ by an average of 0.003A : a mean %age discrepancy of ca. 37%. These results appear adequate for infilling AS=0, and for an approximate subdivision of AS=3. The analysis is continuing, however, with the inclusion of the ratio : N(reflexions) / N(parameters) in the least-squares refinement. These values are not available in CSD, but are being input as an additional regression parameter for a representative sample of entries.

Table Distribution of AS (as %age of Nent) in R-ranges

R-range(%)	AS=	0(%)	1(%)	2(%)	3(%)	4(%)	Nent
1.0-4.0		13	38	32	16	1	8739
4.1-5.0		14	30	32	22	2	7890
5.1-6.0		16	18	32	29	4	6130
6.1-8.0		19	8	27	36	8	7089
8.1-12.0		24	4	16	40	17	3600
overall		16.2	22.5	29.4	26.5	5.3	33448

Figure Plot of $\bar{\sigma}$ vs R for five ranges of Zmax.