MEMORANDUM

23 January 1996

TO: To whom it may concern

FROM: T. Ruf

SUBJECT: Statistical errors in data sets with common events

In studies of systematic effects one is often confronted with the question if measurements using data sets with mainly common events are compatible or not. To answer this question let us consider the following case:

- Sample A with N_A events and measurement $x_A \pm \sigma/\sqrt{N_A}$
- Sample B with N_B events (subset of sample A) and measurement $x_B \pm \sigma/\sqrt{N_B}$
- Sample $\overline{\mathrm{A}}$ with $N_{\overline{A}}$ events (events which are not contained in sample B) and measurement $x_{\overline{A}} \pm \sigma/\sqrt{N_{\overline{A}}}$

We assume that the errors of each measurement scale as $1/\sqrt{N}$. The measurement using sample A is the average of the measurements of sample B and $\overline{\mathrm{A}}$:

$$\begin{aligned} x_A &= \frac{N_B \cdot x_B + N_{\overline{A}} \cdot x_{\overline{A}}}{N_B + N_{\overline{A}}} \\ \sigma_A &= \frac{\sigma}{\sqrt{N_B + N_{\overline{A}}}} \end{aligned}$$

The difference of the measurement x_A and x_B is given by:

$$\begin{aligned} x_A - x_B &= \frac{N_{\overline{A}}}{N_A} \cdot (x_{\overline{A}} - x_A) \\ \sigma_{A-B} &= \sigma_A \sqrt{\frac{N_{\overline{A}}}{N_A} \cdot \frac{1}{1 - \frac{N_{\overline{A}}}{N_A}}} \end{aligned}$$

where x_B and $x_{\overline{A}}$ are uncorrelated measurements.

Example: If sample B contains 80% of the events of sample A then x_A and x_B can still be different as much as $50\% \times \sigma_A$ in 68% of all cases.