BLIND ANALYSIS

P. F. Harrison
Physics Department, Queen Mary University of London, Mile End Rd. London E1 4NS. UK. *

Abstract
We examine what blind analysis is and what motivates its use in particle physics experiments. We explore the methods of blind analysis and give examples of its application in real experiments.

1 WHAT IS A BLIND ANALYSIS?

A blind analysis is an analysis in which the final result, and the individual data on which it is based, are kept hidden from the analyst until the analysis is essentially complete. The principal motivation is to avoid experimenter's (subconcious) bias. It is obvious, but perhaps worth emphasising, that the value of a measurement does not contain any information about its correctness. So knowledge of its value is of no use in performing the analysis itself. The technique has been used in the past by a number of experiments including:

- Many rare decay searches at BNL [1]
- E791 [2]
- KTeV [3]
- *BABAR* [4]
- BELLE [5]

There are several different approaches to blind analysis, the method of choice depending on the type of analysis.

Typically, one has a Monte Carlo or a control sample taken from data to simulate the signal, and signal sidebands in data or a Monte Carlo sample to characterise the background to an analysis. Blind analysis in a counting analysis means optimising all the cuts using such samples, testing fitting procedures, and evaluating the systematic errors before looking at the signal data.

In maximum likelihood fitting, tests may even be made on the signal data, but the fitted values (and any plots which may reveal their approximate values) remain hidden until all checks have been made. Only the errors are output. Some examples of the detailed implementation of blind analysis are given in the later sections of this talk.

2 MOTIVATION FOR BLIND ANALYSIS

The principal motivation is to avoid experimenter's (subconcious) bias. Even the most well-intentioned scientist is susceptible to this! There is a surprising amount of scope for such bias, e.g.

- Looking for bugs when a result does not conform to expectation (and not looking for them when it does).
- Looking for additional sources of systematic uncertainty when a result does not conform
- Deciding whether to publish, or to wait for more data
- Choosing to drop an event which is in the signal region in a rare mode (e.g. because a track is 2 σ away from the particle ID expectation for its type)

The resulting bias represents an unquantifiable systematic uncertainty.

There is evidence in a number of different places in particle physics for such, bias, in addition to well-documented errors in the subject's history. Some possible examples may be found in plots of the

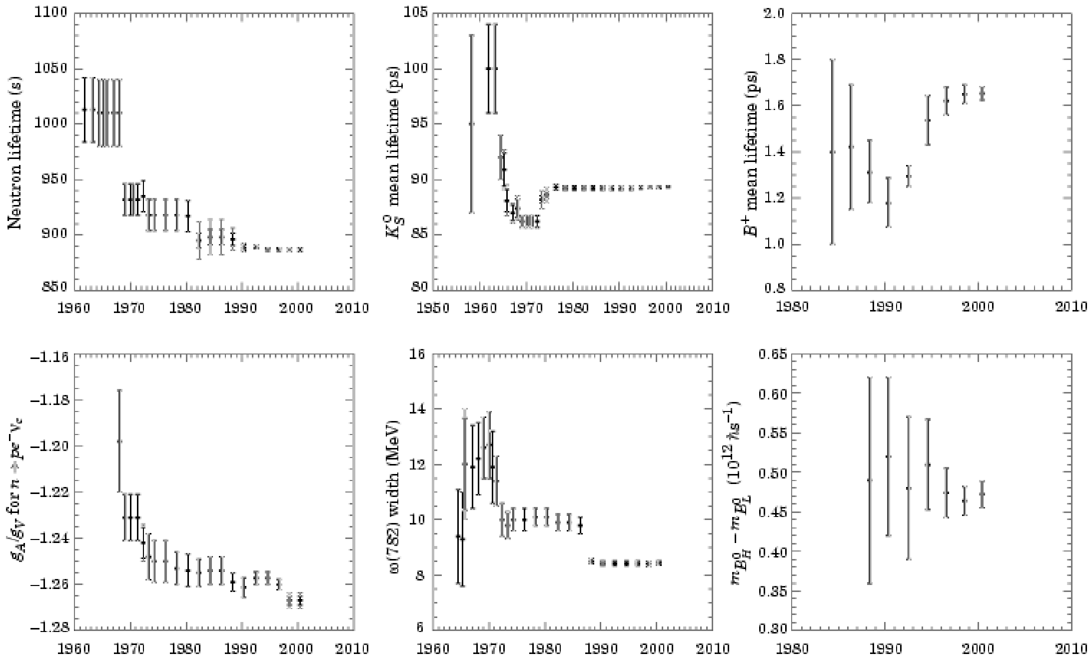

Fig. 1: Some measurements [6] as a function of time

neutron and K_S^0 lifetimes and of the ω width as a function of time, as shown in Fig. 1 [6]. These show periods of surprisingly small variation, followed by jumps of several standard deviations (care should be taken in interpreting the figures, as they actually show the running average, and not just the results of the latest experiments as a function of time).

2.1 Do the LEP Experiments Agree *Too Well*?

Fig. 2 shows the value of R_c as obtained by several of the different LEP experiments. The χ^2 per degree-of-freedom is 0.92/7, showing clearly that there is significantly less variation between the different measurements than one would expect for independent measurements. One possible reason for this could be that perhaps the systematic errors are overestimated, although the original authors would doubtless reject this unlikely interpretation. In order to check this possibility, we also calculated the χ^2 per degree of freedom ignoring the systematic errors and found that it is still only 2.1/7, still rather too small, expecially considering that there must be some real systematic errors, which are ignored in this calculation. Another possible reason for the smallness of the variation between the measurements, in comparison with the quoted errors, is that the measurements are subconciously biased towards each other and/or towards the standard model prediction, perhaps for one or several of the reasons suggested above. We also note that the mean value of all the measurements is surprisingly close to the Standard Model value. One last possible reason for these effects is that we have simply chosen a particularly striking example from an ensemble of ensembles of LEP measurements of different quantities which together display a reasonable distribution of χ^2 values. A complete study of such measurements would surely be an interesting exercise in its own right but is beyond the scope of this talk.

[1]E-mail:p.f.harrison@qmul.ac.uk

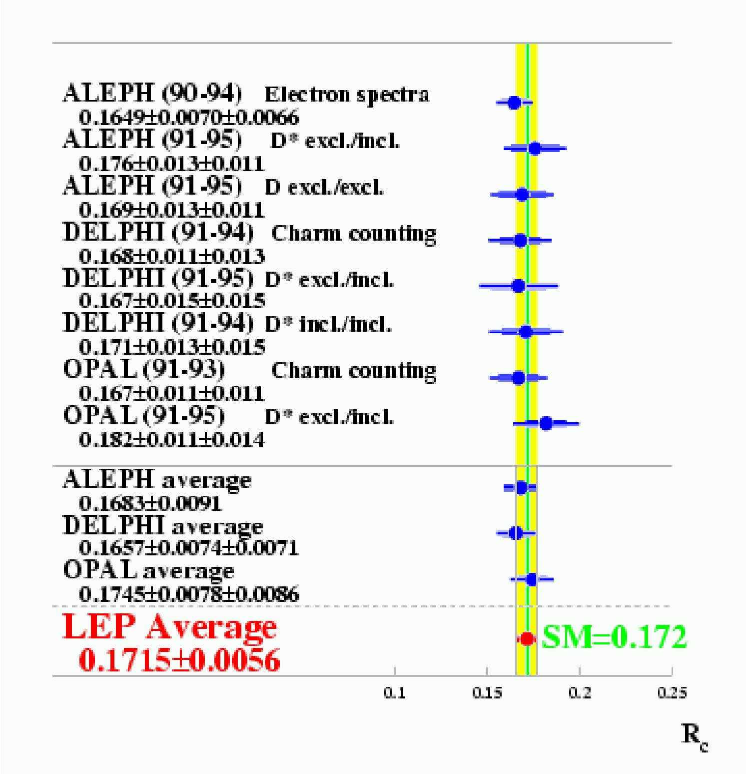

Fig. 2: An ensemble of LEP measurements of the quantity R_c, their average, and the Standard Model prediction for the quantity.

2.2 An Example of Experimenters' Bias: the "Split A2"

At CERN in the mid 1960s, a group using a missing mass spectrometer observed several new mesons in the missing mass spectrum from the process

$$\pi^- + p \to p + MM^-. \tag{1}$$

The A_2 (now known as the I=1 member of the 2^+ nonet) was apparently split, as shown in Fig. 3, and it was fitted with a dipole form. The split A_2 was discussed for several years, and generated considerable speculation by theorists. Similar experiments performed later found no evidence at all for a split. Other experiments gathered data on A_2 via decays to K^+K^- and also found no evidence for a split.

At the Washington APS meeting of 1971, the spokesman of the original CERN experiment, revealed that several cuts which had been made on the data were unneccessary. One of the cuts was based on "running conditions": the group discarded whole runs in which the split did not show up! This is widely regarded as an example of "innocent bias".

2.3 Another Example of Experimenters' Bias: the ζ

In 1984 the Crystal Ball collaboration reported the discovery of a state known as the ζ. It was identified by a peak at $E_\gamma \simeq 1.07$ GeV in the photon spectrum of events reconstructed as:

$$\Upsilon(1S) \to \gamma + X \tag{2}$$

in which X was made up of many hadrons. The "discovery" is shown in Fig. 4.

The ζ would correspond to a particle of mass 8.322 GeV/c^2, width corresponding to the Crystal Ball resolution, and decaying preferentially to multiple hadrons. This would have been an important discovery!

The peak in γ energy was enhanced by the following set of cuts:

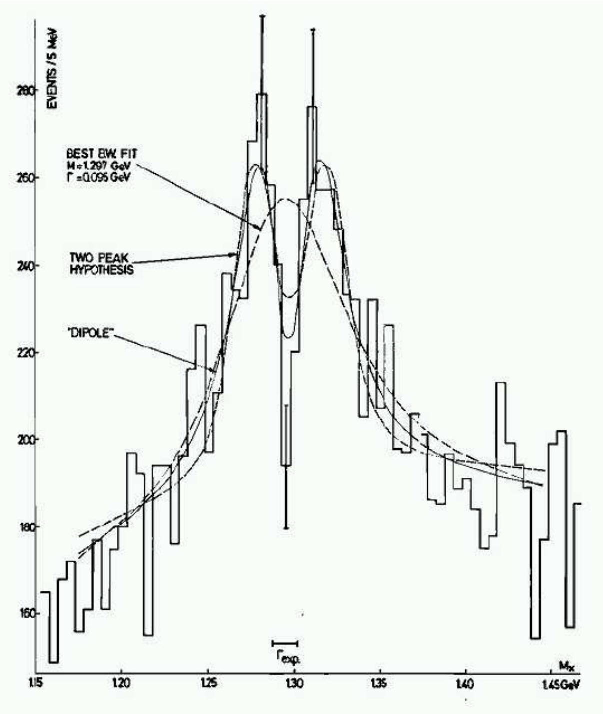

Fig. 3: The "discovery" of the split A_2 in the missing mass spectrum of the process $\pi^- + p \to p + MM^-$.

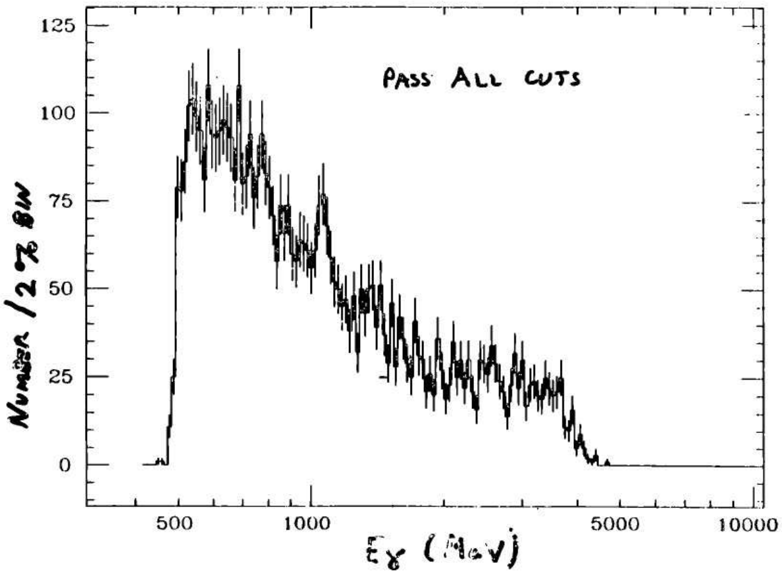

Fig. 4: The "discovery" of the ζ in the photon spectrum of the process $\Upsilon(1S) \to \gamma + X$.

- γ "overlap" (the angle between photon and any hadron) $> 30°$
- $9 <$ Total Multiplicity < 20
- Charged multiplicity ≥ 2
- Neutral multiplicity ≤ 12
- Total energy deposited in calorimeter ≤ 8 GeV
- Sphericity of event ≥ 0.16
- An additional set of "tuning" cuts!

The most effective of these cuts was the first.

However, on removing the γ overlap cut , the photon spectrum changed to that shown in Fig. 5, in which the signal is absent. The problem here was that the cuts were chosen while looking at the data.

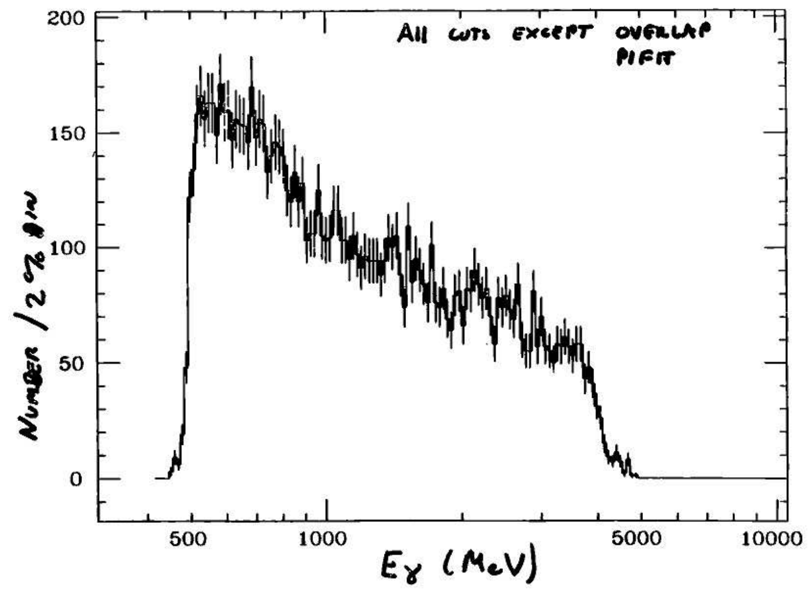

Fig. 5: The same as Fig. 4, but without the γ overlap cut.

Later, the Crystal Ball did a blind analysis on a second batch of data and found that the ζ signal was not real.

As a hint at identifying the kinds of things which may point to a biased analysis, we may note the very arbitrary looking multiplicity cuts, and the "tuning" cuts which seemed to be needed.

3 METHODS AND EXAMPLES OF BLIND ANALYSIS

Blind analysis techniques can be applied to almost any analysis. A few concrete examples include:

- Rare decay searches
- Precision measurements
- CP (rate or time-dependent) asymmetries

We will see how these have been implemented in real cases.

3.1 Blinding in Rare Decay Analyses

By "Rare decays" here, we mean that the branching fraction is not yet measured, or is poorly known. In such cases, the background is probably large, *a priori*, and the analysis must provide a significant background reduction factor. Under these circumstances, a blind analysis is highly desirable!

If the analysis is of the "cut & count" kind, the hidden signal box method is recommended. In this approach, a signal "blinding" box is pre-defined by some cuts, which may be determined using a Monte Carlo signal sample. Blinding means excluding events in the signal blinding box from the analysis AND plots. In practice this can be achieved by implementing a cut which removes the data in the signal region, and (temporarily) filtering the analysis data set with this cut before subsequent analysis steps. The cuts which define the signal blinding box should be slighty looser than those which optimise the analysis, in order to prevent signal in the tails leaking out of the signal region due to poor resolution, or to the fact that the optimal cuts may be quite tight. Sidebands are used to characterise the background in each variable, and the analysis can be optimised using this background characterisation, and signal Monte Carlo. This method assumes that the variables are uncorrelated, so that sidebands in one variable accurately represent the data in the signal region in other variables, an assumption which may be checked with Monte Carlo.

3.2 Example of Rare Decay Blind Analysis: $B^0 \to \rho^{\mp}\pi^{\pm}$ Search at *BABAR*

An example from *BABAR* [7] is shown in Fig. 6, in which the yellow box indicates the signal blinding region. The rectagular box inside this represents the optimised signal box. The shaded regions above, below and to the left of the signal region indicate the various sidebands, in which the shape and normalisation of the background in the two variables can be determined.

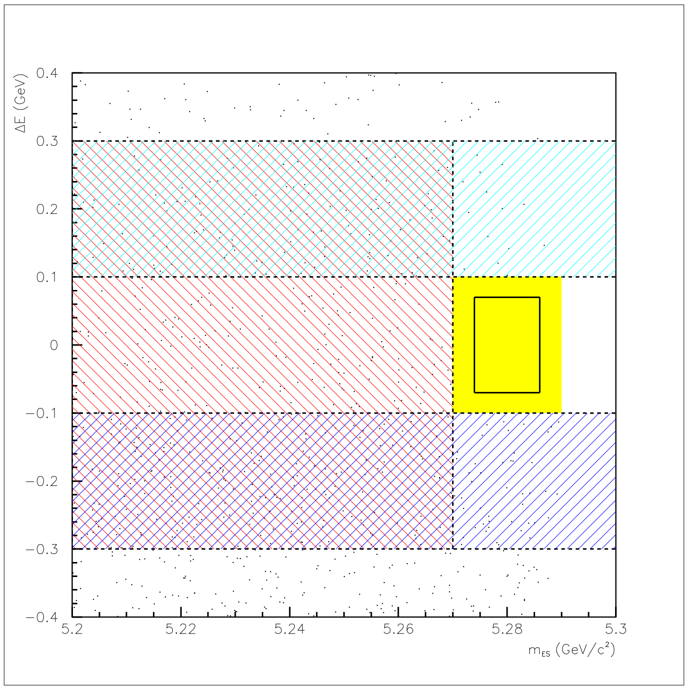

Fig. 6: Scatter plot of the two kinematic variables: $m_{\rm ES} = \sqrt{(\frac{1}{2}s + \mathbf{p}_0 \cdot \mathbf{p}_B)^2/E_0^2 - p_B^2}$, where the subscripts 0 and B refer to the e^+e^- system and the B candidate, respectively; and $\Delta E = E_B^* - \sqrt{s}/2$, where E_B^* is the B candidate energy in the center-of-mass frame. For signal events, the former has a value close to the B meson mass and the latter should be close to zero.

Figures 7 and 8 show respectively the $m_{\rm ES}$ and ΔE distributions for background, determined using on-resonance data in the sidebands of Fig. 6, as well as off-resonance data and Monte Carlo.

3.21 Blind Cut Optimisation with (1/2 of) the Data

The final sample is defined by cuts in many discriminating variables, including $m_{\rm ES}$ and ΔE. The cuts are optimised with respect to some objective figure of merit e.g. the statistical significance (number of

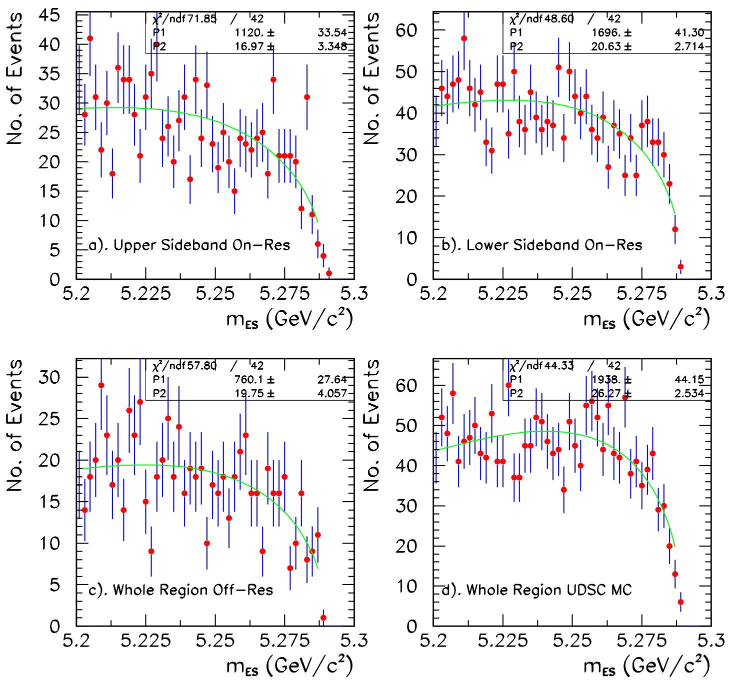

Fig. 7: $m_{\rm ES}$ distribution for background, determined using on-resonance data in a) the upper sideband of Fig. 6, b) the lower sideband, c) off-resonance data and d) background Monte Carlo.

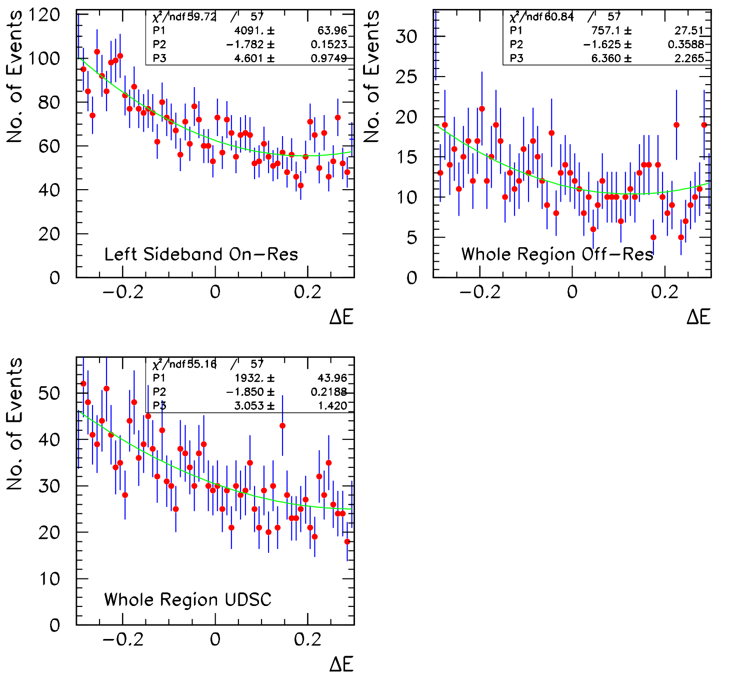

Fig. 8: ΔE distribution for background, determined using on-resonance data in a) the left sideband of Fig. 6, b) off-resonance data and c) background Monte Carlo.

standard deviations):

$$\frac{S}{\sqrt{(S+B)}} \tag{3}$$

where S and B are the (expected) numbers of signal and background events in the final cut-optimised samples. S and B are obtained (as functions of the cut values) from signal Monte Carlo or control samples for signal (efficiency) and from sidebands for background, while the actual signal region remains blind. It is important to avoid over-tuning cuts into statistical fluctuations in these samples, but large sample sizes help (the sidebands can be somewhat larger than the eventual signal region, which helps to increase the statistics available for this optimisation).

In order to avoid biasing the efficiency, only e.g. 1/2 of the data (sideband and Monte Carlo) are used in the cut-optimisation. Once the cuts have been defined, the other half of the data and Monte Carlo samples are used to obtain the background normalisation and the signal efficiency.

3.22 Unblinding

In *BABAR* an analysis will normally have been presented to an Analysis Working Group (AWG) before unblinding. The presentation will include:

- A description of the cut optimisation and background characterisation procedures.
- The expected number of background events in the signal box (as a function of the branching fraction, if this is completely unknown)
- The signal efficiency from Monte Carlo or control samples
- The expected statistical sensitivity
- An estimation of systematic errors

After discussion, permission is sought from the AWG to unblind the signal. For a rare decay cut and count analysis, unblinding is essentially just a counting exercise: how many events are inside the signal box?

After unblinding, plots showing the signal region can be made. Depending on the importance of the analysis, the date and time of unblinding may be publicised widely within the collaboration. For a particularly high profile analysis, there may even be an “unblinding party” in a pubic terminal area, at which the crucial piece of code is run to “open the box”, ie. to reveal the data in the signal region.

Finally, the appropriate plots showing the unblinded signal region can be made. Figure 9 shows the signal found in the *BABAR* $B^0 \to \rho^{\mp}\pi^{\pm}$ branching fraction analysis, as an excess of events above the background.

3.3 Blinding in Precision Measurements

Precision measurements, such as Δm_d or B and D lifetimes are good candidates for blind analyses. In such cases, accurate prior measurements generally exist and use of a blind analysis removes the possibility of a bias towards the PDG values. The analysis is probably systematics-limited, and blind analysis ensures that the choices involved in estimating the systematic uncertainties are not biased by the value of the result.

In such cases, the method often involves a maximum likelihood fit. Then, systematic checks necessarily involve re-fitting on the data. In order to blind the measurement while performing such checks, one can use the hidden offset method.

3.31 The Hidden Offset Method

In this approach, the fitting code adds a fixed, unknown pseudo-random number (or numbers), $\mathcal{R}$, to the fitted value of the measured parameter(s):

$$x^* = x + \mathcal{R}. \tag{4}$$

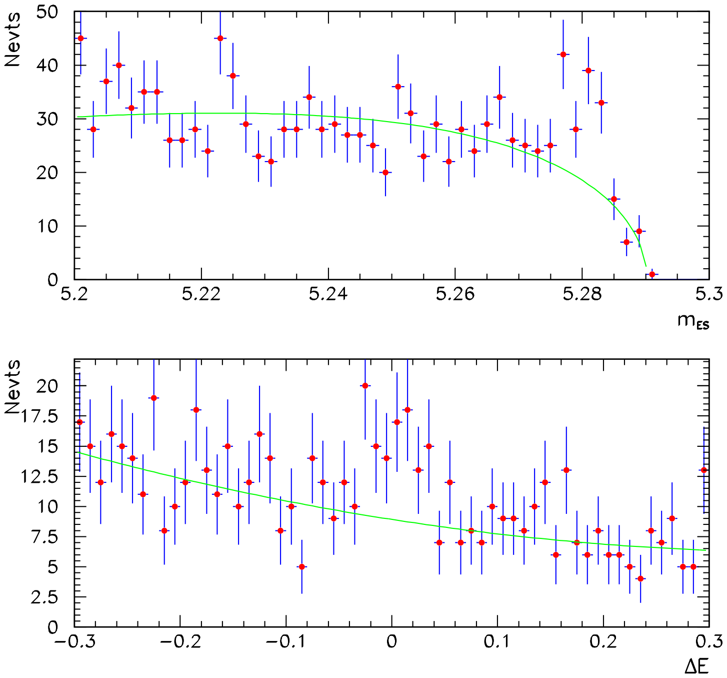

Fig. 9: The $B^0 \rightarrow \rho^{\mp}\pi^{\pm}$ signal region projected onto m_{ES} and ΔE axes respectively. The signal can be seen as an excess of events above the background level which is indicated by the curve shown.

x^* is returned (along with the true error and likelihood value) instead of x.

$\mathcal{R}$ is sampled from a Gaussian distribution with mean 0 and standard deviation equal to a few times the experimental standard deviation. Any plot of the likelihood function should also have the random offset applied to the values of the fitted variable, so that the fitted value cannot be read-off from the minimum. Relative changes in the result, as changes are made in the analysis (new decay modes added to the sample, for example) can be hidden using a second offset:

$$x^* = 2x_{PDG} - x + \mathcal{R}. \tag{5}$$

One of the the two methods is chosen at random after each change in the analysis.

3.4 Blinding ϵ'/ϵ at KTeV

KTeV [3] is a precision counting experiment - ϵ'/ϵ depends on the numbers of K_S^0 and K_L^0 decays each to $\pi^+\pi^-$ and $\pi^0\pi^0$. The experimenters wanted to use all the events to make systematic checks (measuring $\tau_{K_S^0}$ and Δm_K) and to check their method of extraction of ϵ'/ϵ, without biasing themselves. In other words, they wanted to look at their data without being biased. The blinding strategy which was adopted allowed this. They performed the analysis just as any other experiment, except that the value of ϵ'/ϵ was kept hidden from the experimenters until the analysis was complete. They fitted the data and hid the value of ϵ'/ϵ as follows:

$$[\epsilon'/\epsilon]^* = \left\{ \begin{array}{c} 1 \\ -1 \end{array} \right\} \times \epsilon'/\epsilon + \mathcal{R}. \tag{6}$$

The $\{1$ or -1$\}$ and $\mathcal{R}$ were chosen at random but kept fixed, once chosen. The sign flip was to make it uncertain whether the value of ϵ'/ϵ was going up or down with analysis changes.

Toy Monte Carlo fits were done with no blinding to verify their fitting procedures. Two analyses were performed in parallel, with different random choices, to avoid biasing each other. Once both were completed, the random choices were made the same: a blind comparison could be made! The result was finally unblinded and announced to the public one week later. No analysis changes were made during this week.

3.5 Blinding Time-dependent CP Asymmetries

The time-dependent CP asymmetry in neutral B meson decays to CP-eigenstates is manifested in the shapes of the time- ($\sim \Delta z$-)distributions of such decays, for both B^0 or $\overline{B}^0$ mesons. It is fit by the maximum likelihood method using the hidden offset method. However, the asymmetry is readily visible by eye, even with small statistics, in both distributions separately, as well as by comparing them, as can be seen in Fig. 10. This visual clue also needs to be blinded.

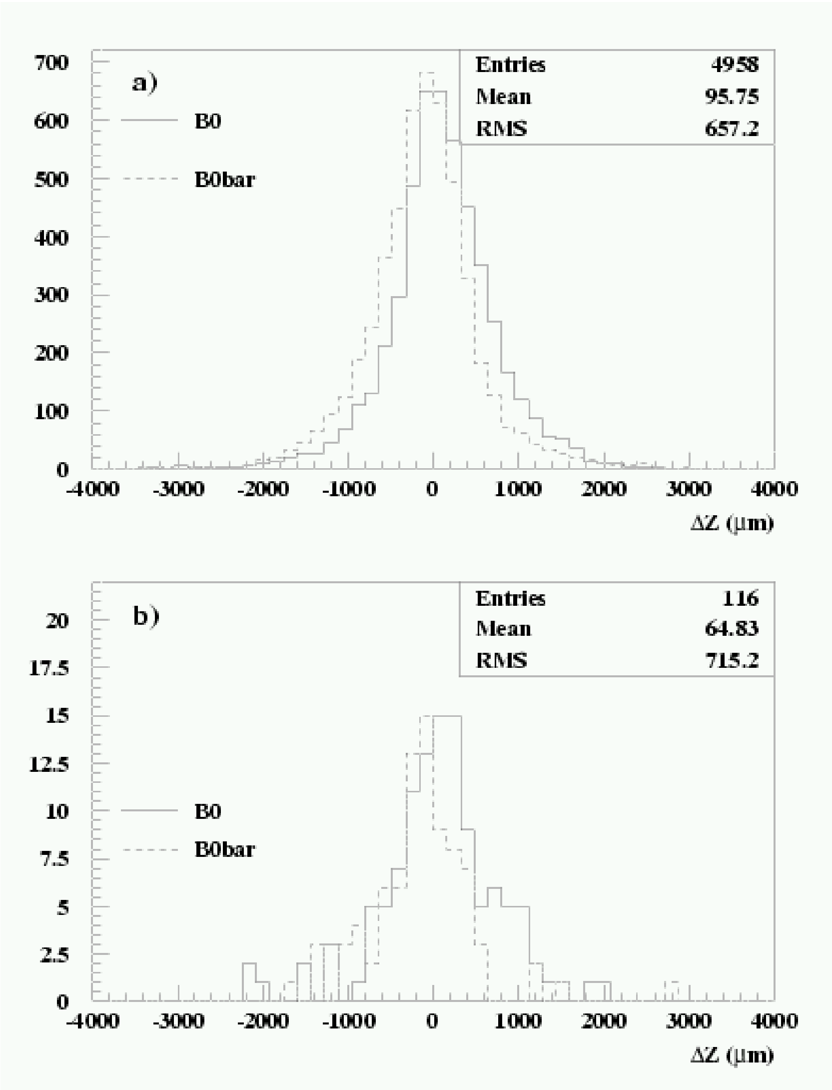

Fig. 10: Time-dependent CP asymmetries in the golden mode $B^0 \rightarrow J/\psi K^0_S$, for a) high statistics and b) low statistics samples.

This can be achieved by plotting only

$$[s_{tag}\Delta z]^* = s_{tag}\Delta z + \mathcal{R} \tag{7}$$

where s_{tag} is the sign of the tag. No information is lost, except the asymmetry between B^0 and $\overline{B}^0$, and the magnitude of the asymmetry between positive and negative Δz. Example plots are shown in Fig. 11, in which the histograms for B^0 and $\overline{B}^0$ are approximately congruent, and the magnitude and sign of the asymmetry has been hidden by the unknown random variable $\mathcal{R}$. Hiding the fitted asymmetry is a

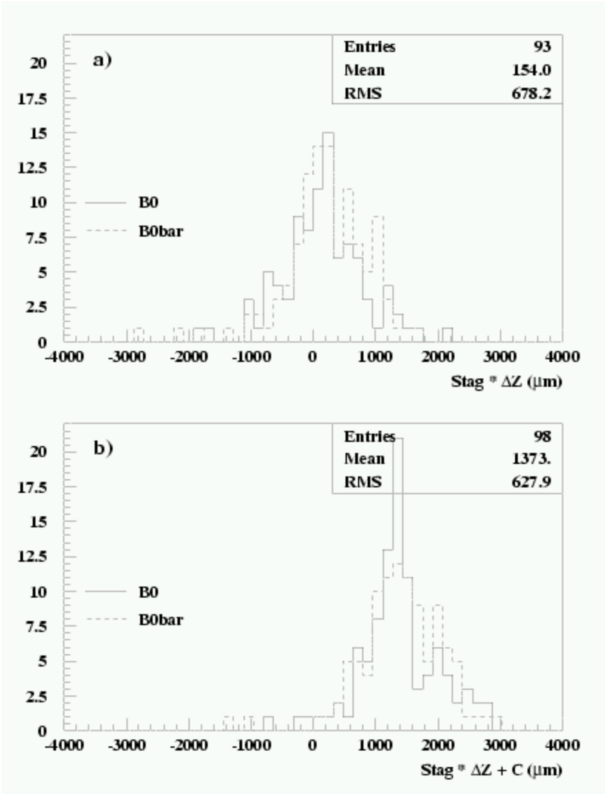

Fig. 11: Time-dependent CP asymmetries in the golden mode $B^0 \to J/\psi K^0_S$ with the tag information hidden. In a), the relative sign between B^0 and $\overline{B}^0$ has been removed by multiplication by the factor s_{tag}. In b) the random offset (with random sign) has also been added.

separate operation, exactly analogous to hiding ϵ'/ϵ.

4 CLOSING COMMENTS

Blind analysis brings particle physics into line with best practice from other branches of science. It is more a formalisation of good experimental practice than a radical new idea. It is certainly no panacea, and is not a substitute for careful, thoughtful analysis, but it is an additional safeguard. An analysis which is not blind is not necessarily a wrong analysis, and an analysis which is blind is not necessarily a right analysis. However, the field has had its fair share of embarrassing wrong results and the technique can only help in reducing these in the future. Even when an unblind analysis has been performed in an unbiased way, just the possibility of experimenters' bias reduces our own and others' confidence in our results. If we can reduce the risk of bias, why not do so?

Acknowledgements
I would like to acknowlege the significant contributions of my collaborators: Brian Meadows, Jeff Richman, Aaron Roodman and Alan Watson, from whom I learned a great deal about blind analysis, and the dangers of not performing one, and from whose own talks and papers on the subject I have borrowed shamelessly in writing this talk.

References

[1] K. Arlsala et al. Phys. Rev. Lett. 70 (1993) 1049;
K. Arlsala et al. Phys. Rev. Lett. 71 (1993) 3910;
S. Adler et al. Phys. Rev. Lett. 76 (1996) 1421;
J. Belz et al. Phys. Rev. Lett. 76 (1996) 3277.

[2] E. Altala et al. Phys. Rev. Lett. 81 (1998) 44;
E. Altala et al. Phys. Lett. B448 (1999) 303;
E. Altala et al. Phys. Lett. B462 (1999) 401

[3] A. Alavi-Harati et al. Phys. Rev. Lett. 83 (1999) 22.

[4] B. Aubert et al. (*BABAR* Colaboration), hep-ex/0201020.

[5] K. Abe et al. (BELLE Colaboration), hep-ex/0202027, Submitted to Phys. Rev. D.

[6] D.E. Groom et al. Eur. Phys. J. C15 (2000) 1.

[7] B. Aubert et al. (*BABAR* Colaboration), hep-ex/0107058.