Blind Analysis

Paul Harrison

Queen Mary University of London

Durham Workshop on Advanced Statistical Techniques

in Particle Physics

Acknowledgements: B. Meadows, J. Richman, A. Roodman, A. Watson

Outline of Talk

Things I will talk about:

- What is Blind Analysis?
- Motivation for Blind Analysis
- Applications, Methods and Examples
- Concluding Remarks

Bias and Experiment

The value of a measurement does not contain any information about its correctness!

"It's OK to be stupid, but it's not OK to be biased!"

- Ed Thorndike, Panofsky Prize Winner, 1999

ie. it's not that important to be optimal, but it is essential to be unbiased

What is a Blind Analysis?

It is an analysis in which the final result (and the individual data which contribute to it) is kept hidden from the analyst until the analysis is essentially complete.

Originally proposed in HEP context by W.R. Molzon, the concept was born in the field of rare decay searches.

Has been used in the past by a number of experiments including:

- Many rare decay searches at BNL
- E791
- NA31
- KTeV
- BABAR
- BELLE

What is a Blind Analysis?

Different approaches to BA - depends on the type of analysis.

Typically, one has MC or control sample to simulate signal and sidebands or MC to characterise the BG.

BA in counting analysis means optimising all the cuts using such samples, testing fitting procedures, and evaluating systematic errors before looking at the signal data.

In ML fitting, tests may even be made on the signal data, but the fitted values (and plots which reveal them) remain hidden until all checks have been made. Only the errors are output.

Some examples of detailed implementation later.

Motivation for Blind Analysis

Principal motivation is to avoid experimenter's (subconcious) bias.

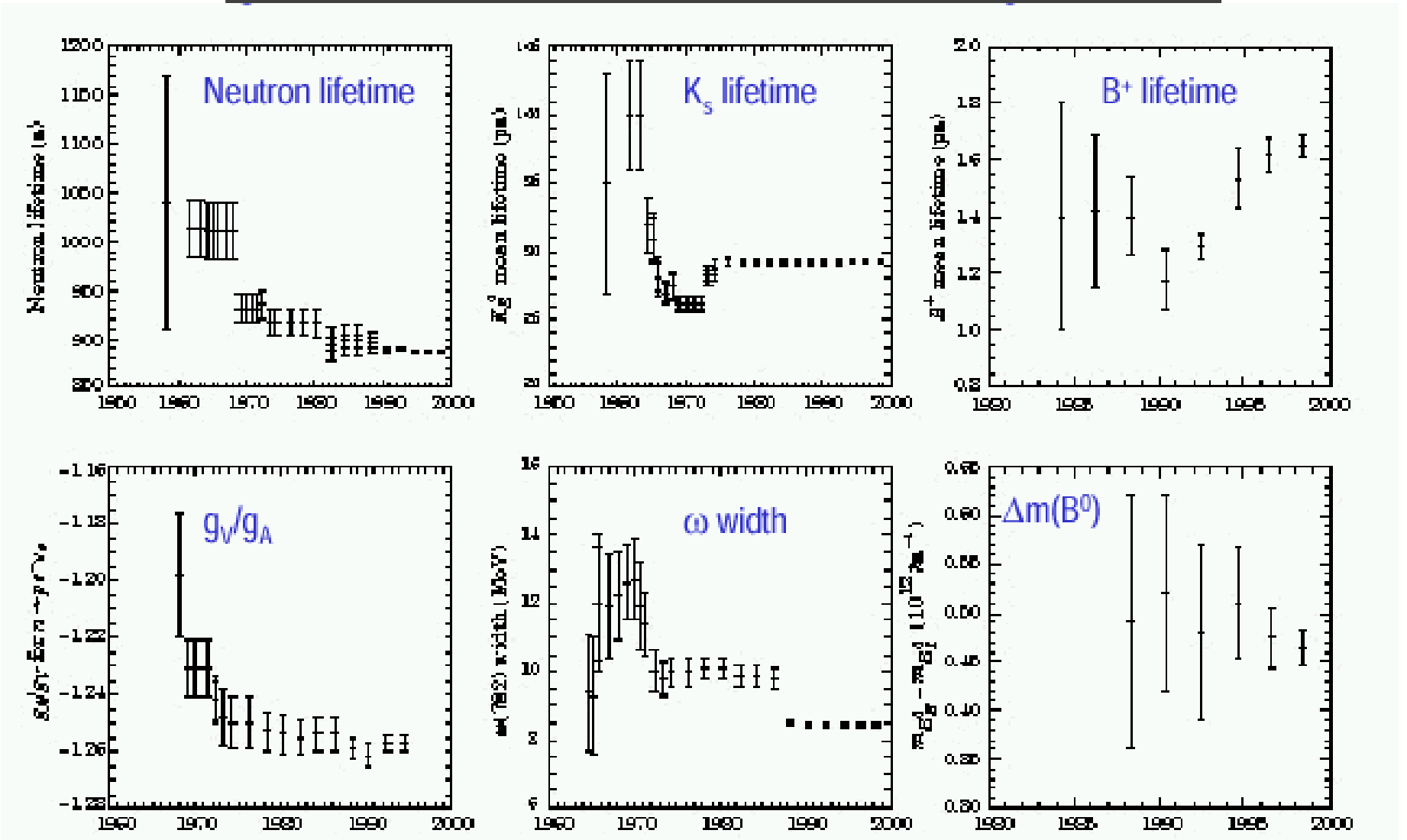
Even the most well-intentioned scientist is susceptible to this!

Surprising amount of scope for such bias:

- Looking for bugs when a result does not conform to expectation (and not looking for them when it does).
- Looking for additional systematic errors 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 (eg. because a track is 2 σ away from the particle ID expectation for its type)

Resulting bias represents an unquantifiable systematic error.

Some Measurements as a function of Time I



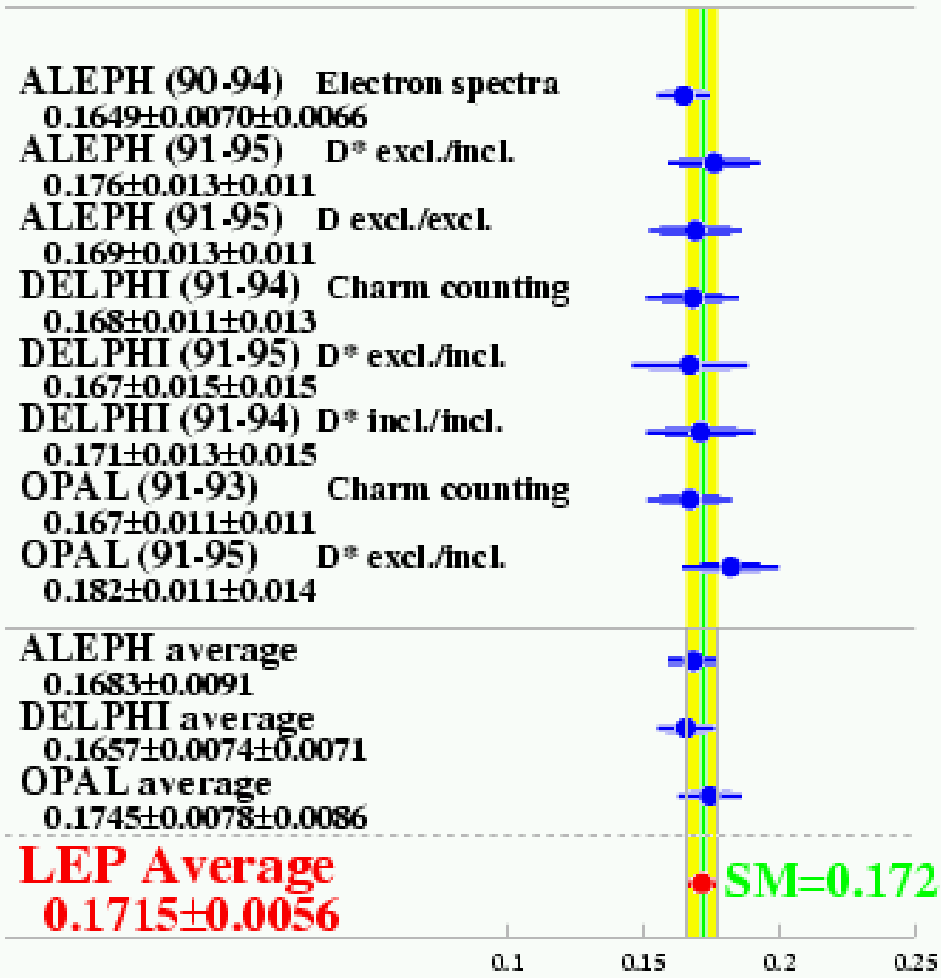
Plots of neutron and K_s^0 lifetime and ω width show periods of surprisingly small variation, followed by jumps of several standard deviations.

Do the LEP Experiments Agree Too Well?

Yes! (agreement with SM is also v. good cf. error.)

Is this because:

- experimenters' bias has pulled the results closer to each other and to SM value?
- systematic uncertainties are overestimated?
- Or have we simply chosen one result which happens to have fluctuated "down" in χ^2?



$\chi^2/\mathrm{dof} = 0.92/7$

(=2.1/7 ignoring systematic errors!).

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

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

$$\pi^- p \to p + MM^-$$

- The A_2 (now known as the I=1 member of the 2^+ nonet) was apparently split. It was fitted with a dipole form.
- The split A_2 was discussed for several years, and generated considerable speculation by theorists on causality, non-local field theory etc.

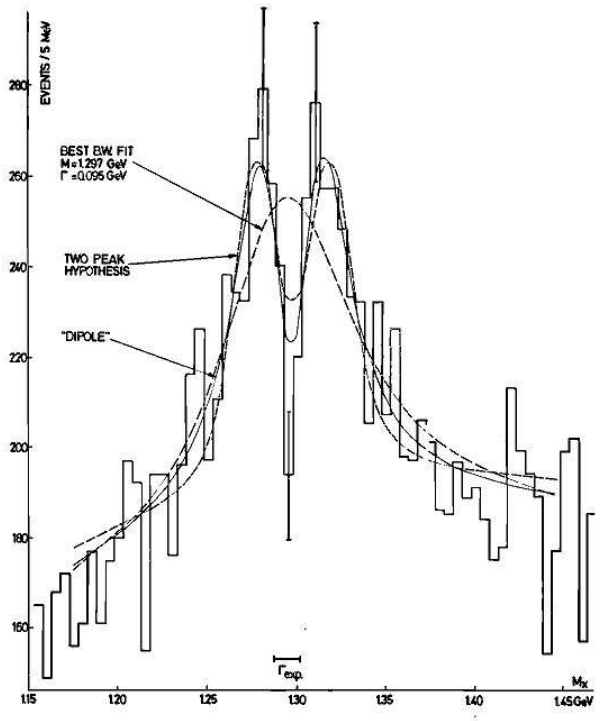


Fig. 4. Total A_2. (May 1965 + October 1965 + January 1967). The following fits were obtained:
1) a single Breit-Wigner curve (dash-point line);
2) two incoherent Breit-Wigner curves (dashed line);
3) a "dipole" (full line).

The Split "A2" Contd.

- Similar experiments performed later found no evidence at all for a split.
- Other experiments gathered data on A_2 via decays to K^+K^-. They found no evidence for a split.
- At the Washington APS meeting of 1971, B. Maglic, 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". His group discarded whole runs in which the split did not show up!
- This was widely regarded as an example of "innocent bias".

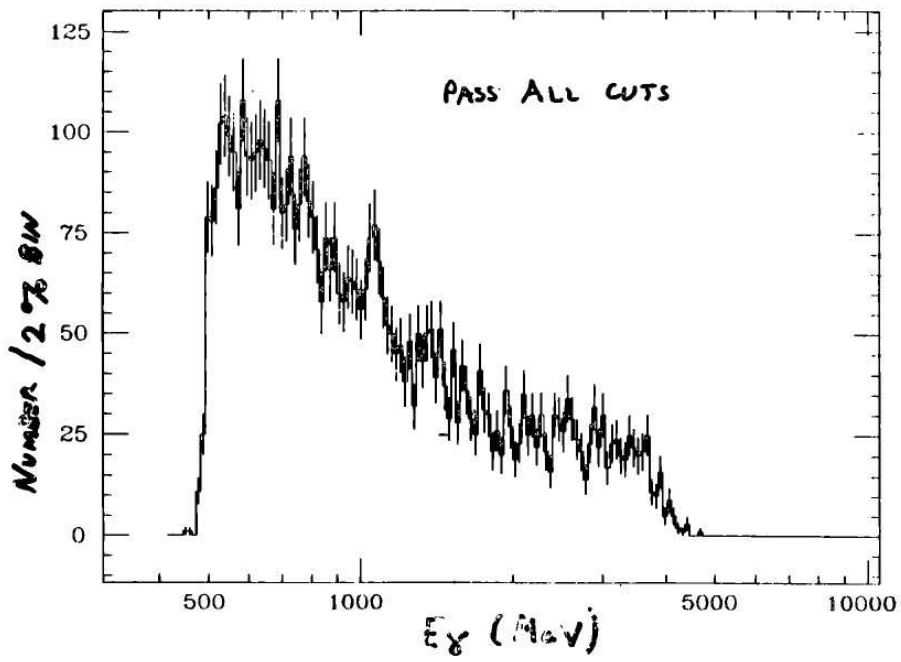
Another Example of Experimenters' Bias: the ζ

- In 1984 the Crystal Ball Collaboration reported the discovery of a state known as the ζ. It was identified in the photon spectrum of events reconstructed as:

$$\Upsilon(1S) \rightarrow \gamma + X$$

in which X was made up of many hadrons. Note the peak at $E_\gamma \simeq 1.07$ GeV.

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



The ζ contd.

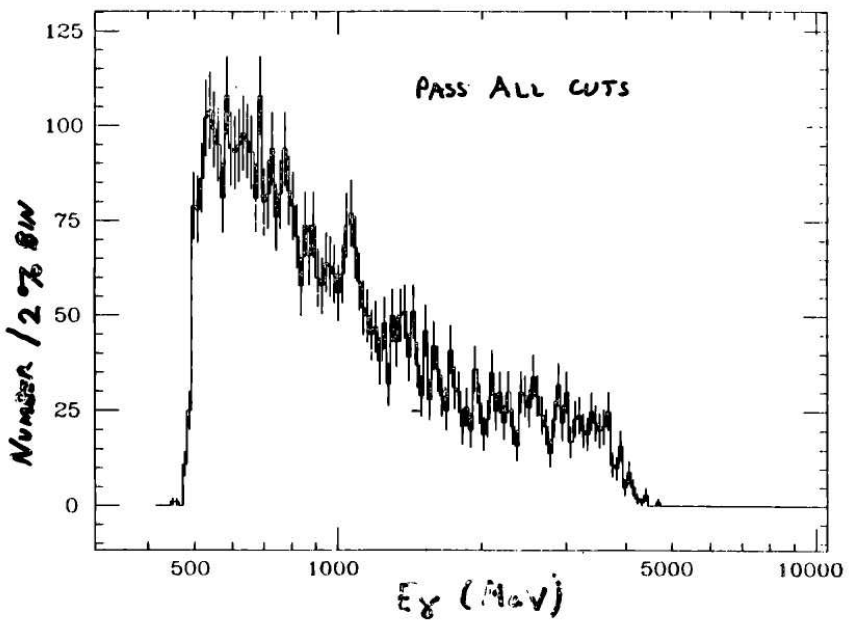
The peak in γ energy was enhanced by the following set of cuts:

- 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 addition set of "tuning" cuts!

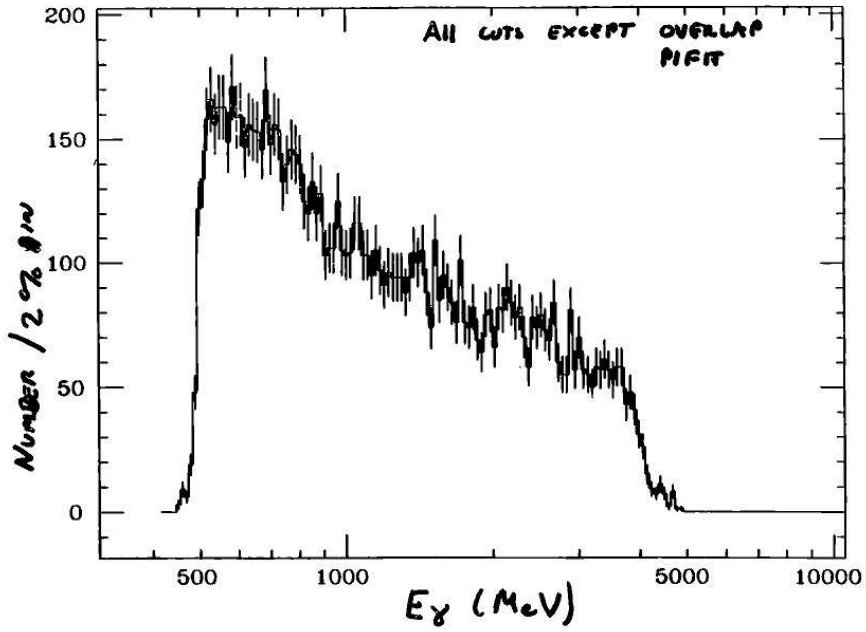
The most effective of these cuts was the first.

The ζ contd.

With γ overlap cut:



Without γ overlap cut:



What went wrong? The cuts were chosen while looking at the data. Note:

- the arbitrary multiplicity cuts
- the "tuning" cuts needed?

Later, the CB did a Blind Analysis on a second batch of data and found that the ζ signal was not real.

Applications and Methods of Blind Analysis

The BA technique 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.

Blinding in Rare Decay Analyses

"Rare decays" $\Longrightarrow$

- BF not yet measured, or poorly known
- Backgrounds probably large: analysis must provide significant reduction factor

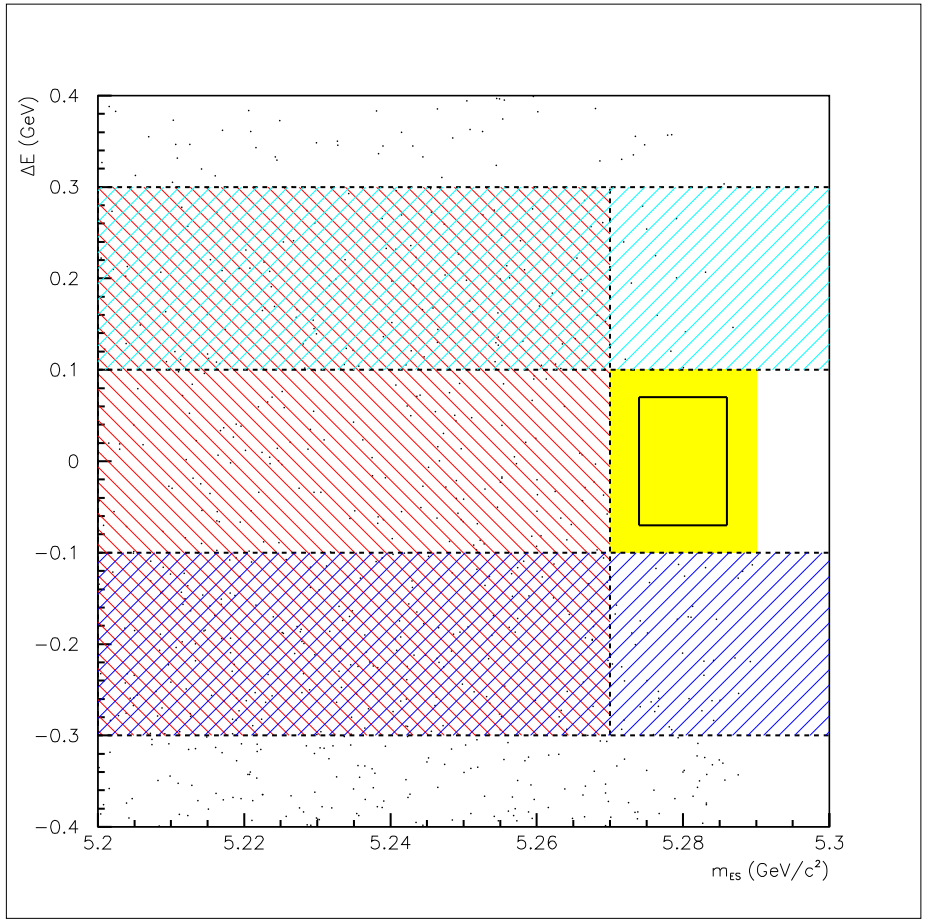
Then blind analysis highly desirable!
For "cut & count" analysis, the hidden "signal box" method is recommended.

Signal box defined by loose cuts.

Blinding means excluding events in signal box from analysis AND plots.

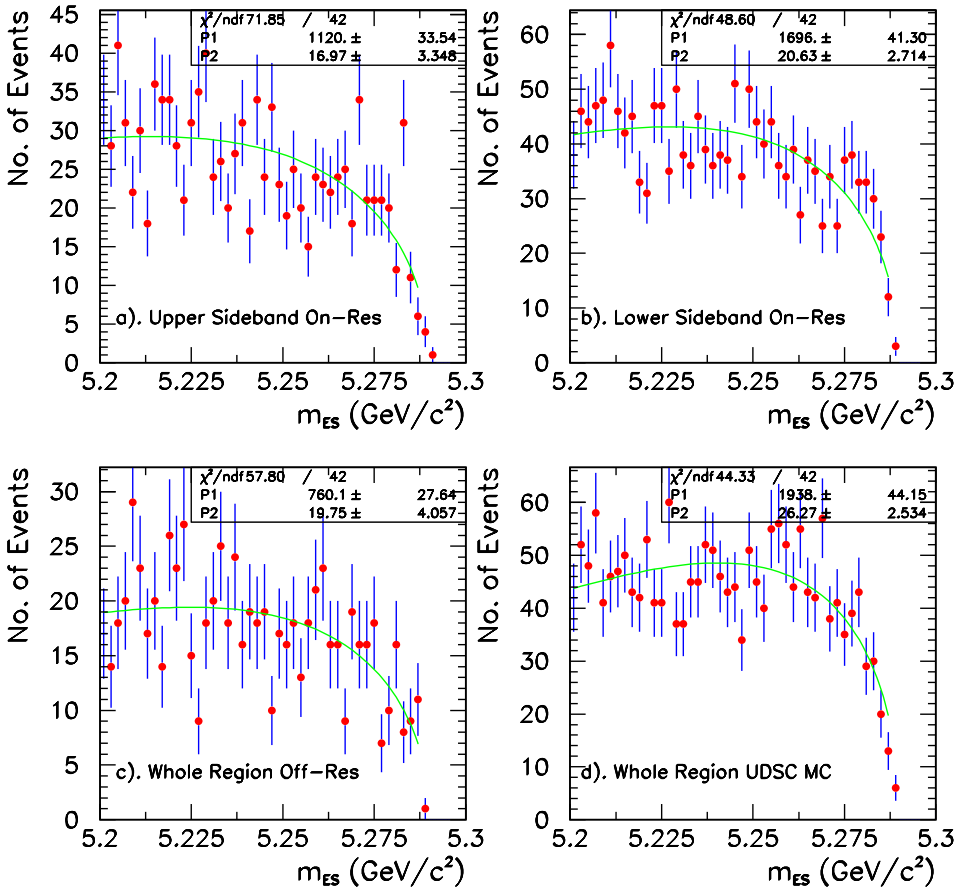
Sidebands are used to characterise the background in each variable

Assumes that the variables are uncorrelated - checked in Monte Carlo.

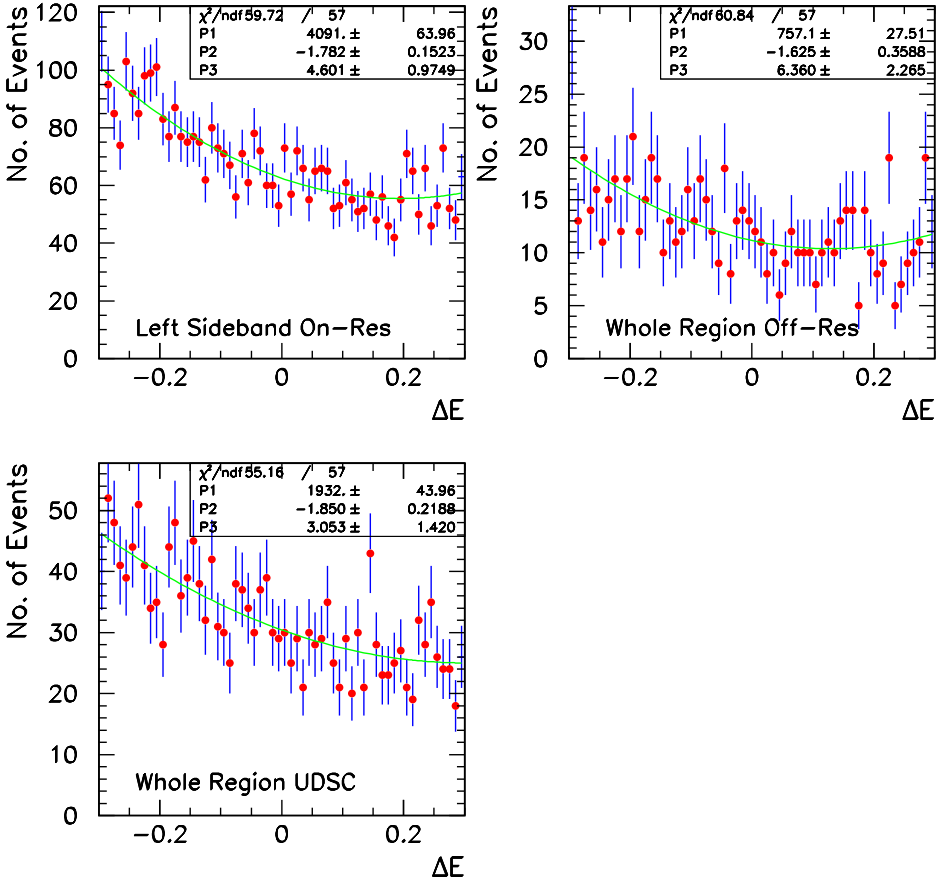


Blind Background Characterisation in $B^0 \to \rho^{\mp}\pi^{\pm}$ Search at BABAR

BG $m_{\rm ES}$ distribution:



BG ΔE distribution:



Absolute level of background in signal region is easily calculated (later) by extrapolating BG to the (hidden) signal box.

Blind Cut Optimisation with (1/2 of) BABAR Data

- Final sample is defined by cuts in many discriminating variables.
- Cuts optimised with respect to some objective figure of merit eg. the statistical significance (# of σ):

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

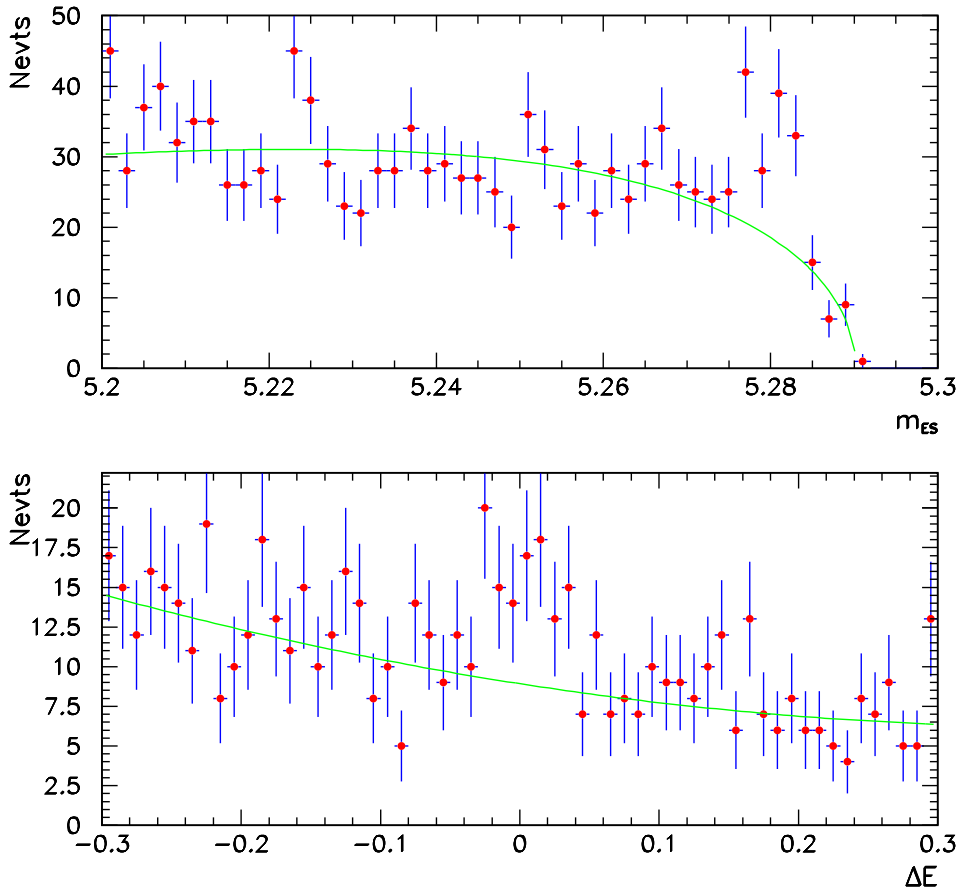
  where S and B are the numbers of signal and BG events in the final signal.
- S and B obtained (as function of cut values) from signal MC or control samples for signal (efficiency) and from sidebands for BG.
- eg. For $B^0 \to \rho^{\mp}\pi^{\pm}$, used $B^+ \to \rho^+ D^0(\to K^+\pi^-)$, to validate signal MC.
- Need to avoid over-tuning cuts into statistical fluctuations in these samples, but large sample sizes help.
- To avoid residual efficiency bias, use only eg. 1/2 of data (sideband and MC). Once cuts defined, use other 1/2 to normalise BG and obtain MC efficiencies.

Unblinding

- In BABAR an analysis will normally have been presented to an Analysis Working Group before unblinding. The presentation will include:
  - Description of cut optimisation and BG characterisation.
  - Expected # of background events in signal box
  - Signal efficiency from MC or control samples
  - Expected statistical sensitivity
  - Estimation of systematic errors
- After discussion, it can be unblinded.
- For a rare decay mode, this is essentially just a counting exercise: how many events are inside the signal box?

Unblinding (contd.)

- 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 experiment.
- For sin2β, there is an "unblinding party" in a pubic terminal area, at which the crucial piece of code is run to "open the box".



The $B^0 \to \rho^{\mp}\pi^{\pm}$ signal projected onto $m_{\rm ES}$ and ΔE axes respectively.

Blinding in Precision Measurements

Precision measurements, such as Δm_d or B and D lifetimes are good candidates for blind analyses

- Generally very accurate prior measurements exist
- Use of BA removes possibility of bias towards PDG values
- Probably systematics-limited, and BA ensures that the choices involved in estimating these are not biased by the value of the result
- Method probably involves a maximum likelihood fit

When ML fitting is involved, 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.

The Hidden Offset Method

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

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

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 $\sigma \sim$ few σ_{exp}.

A plot of the likelihood function should also have the random offset.

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:

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

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

Blinding ϵ'/ϵ at KTeV

A precision counting experiment - ϵ'/ϵ depends on the numbers of K_S^0 and K_L^0 decays each to $\pi^+\pi^-$ and $\pi^0\pi^0$.

- Needed all events to make systematic checks (measuring $\tau_{K_S^0}$ and Δm_K).
- Needed to check their extraction of ϵ'/ϵ, without biasing themselves.

Blinding strategy allowed these and more.

Was just like any other experiment except that the value of ϵ'/ϵ was kept hidden

They fitted the data and hid the value of ϵ'/ϵ as follows:

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

The {1 or -1} and $\mathcal{R}$ were chosen at random and kept fixed.

Sign flip was to make it uncertain whether the value was going up or down with analysis changes.

Unblinding ϵ'/ϵ at KTeV

Toy MC fits were done with no blinding to verify fitting procedures.

Two analyses were performed in parallel with different random choices, to avoid biasing each other.

When completed, the random choices were made the same: a blind comparison could be made!

Result was finally unblinded and announced to the public one week later.

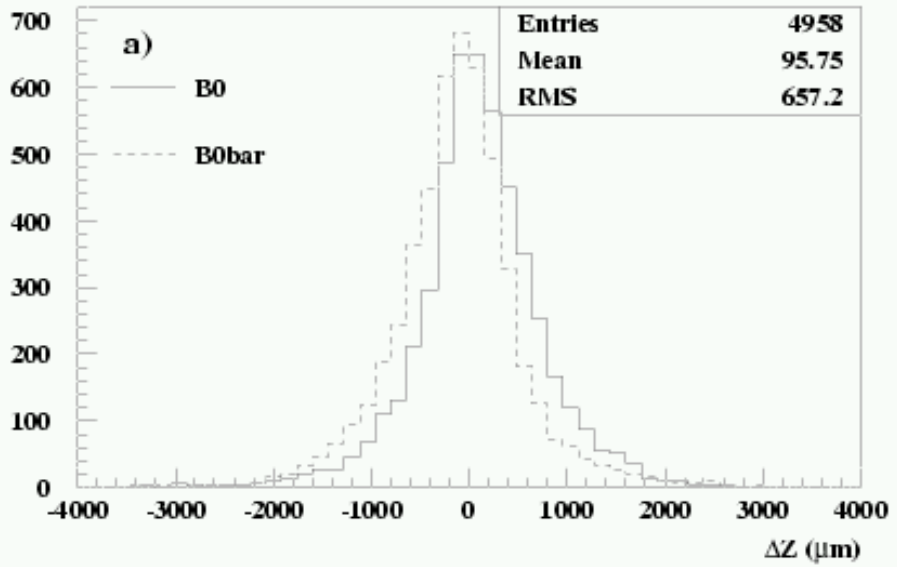
No analysis changes were made during this week.

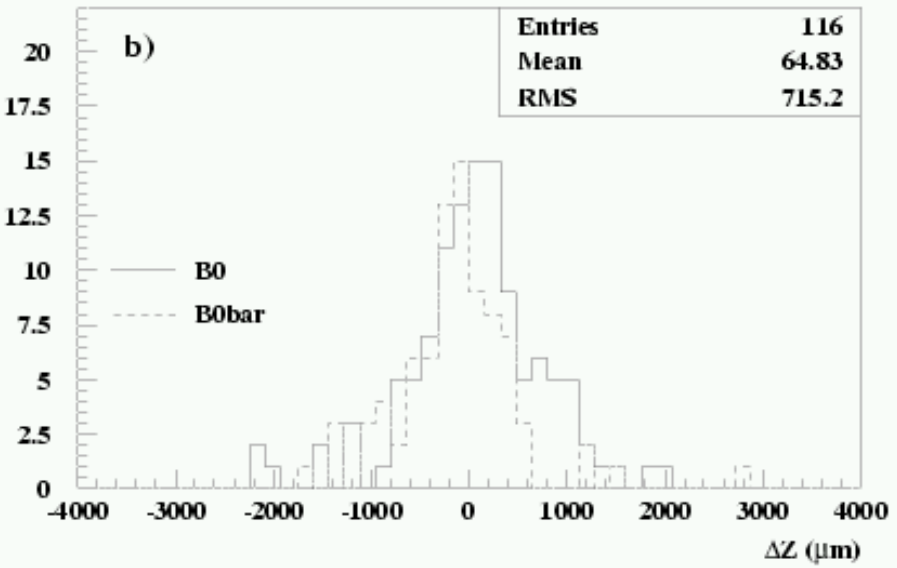
Time-dependent CP Asymmetries

The time-dependent CP asymmetry in neutral B decays to CP-eigenstates can be seen easily in the shapes of the time-($\sim \Delta z$)-distributions of such decays of either B^0 or $\overline{B}^0$.

Can fit by ML using hidden offset method.

But asymm is readily visible, even with small stats, in both distributions separately, as well as by comparing them. This visual clue needs also to be blinded.





Blinding Time-dependent CP Asymmetries

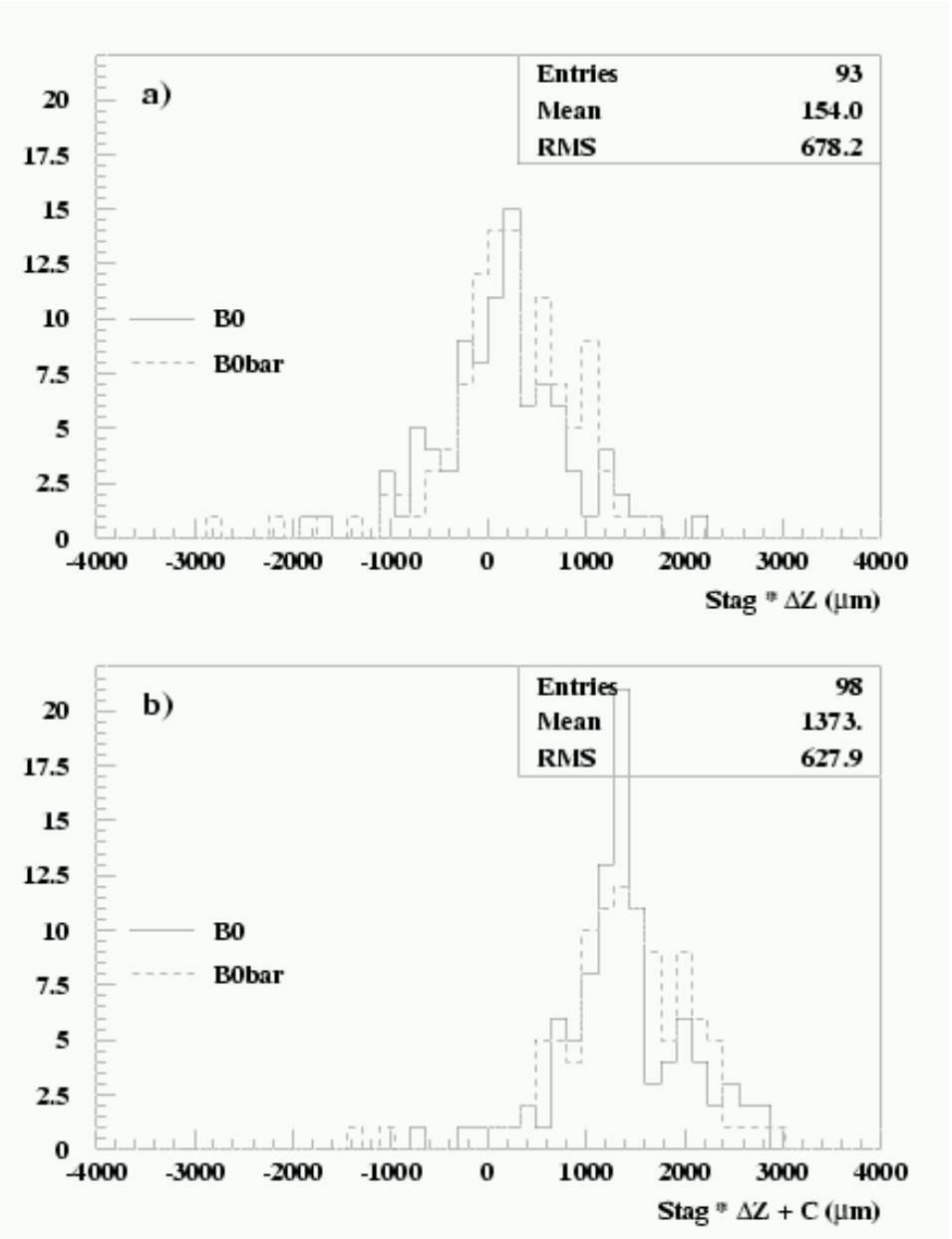
Here, can avoid seeing the asymmetry in either distribution by plotting only

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

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 +ve and -ve Δz.

Hiding the fitted asymmetry is a separate operation, exactly analogous to hiding ϵ'/ϵ.



Closing Comments

- BA brings particle physics into line with best practice from other branches of science.
- More a formalisation of good experimental practice, than a radical new idea.
- An analysis which is not blind, is not necessarily a wrong analysis
- An analysis which is blind is not necessarily a right analysis
- The field has its fair-share of embarrassing wrong results
- Even the chance of experimenters' bias reduces our confidence in our results.
- If we can reduce risks of bias, why not do so?