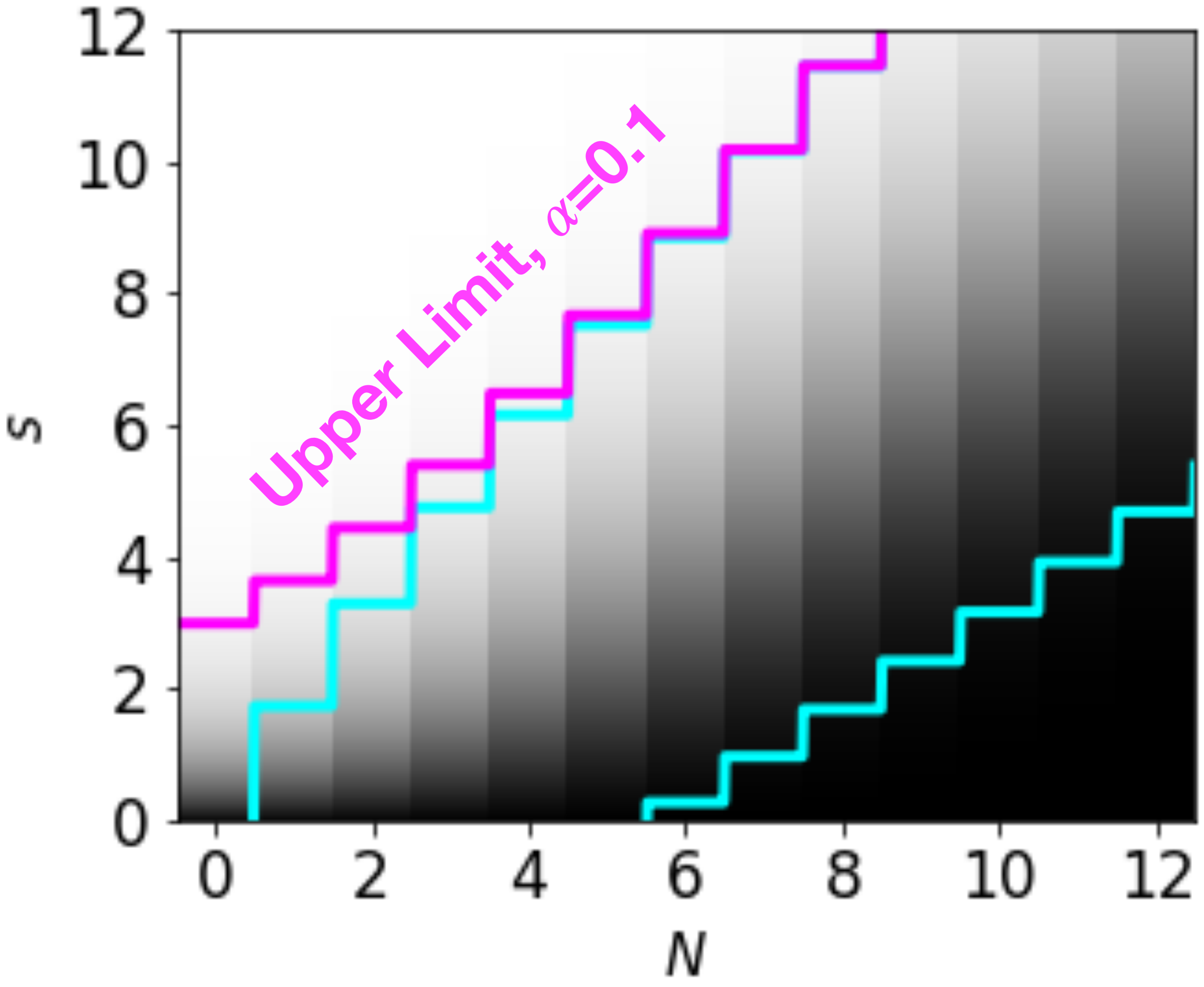
Imposing overcoverage on small signals

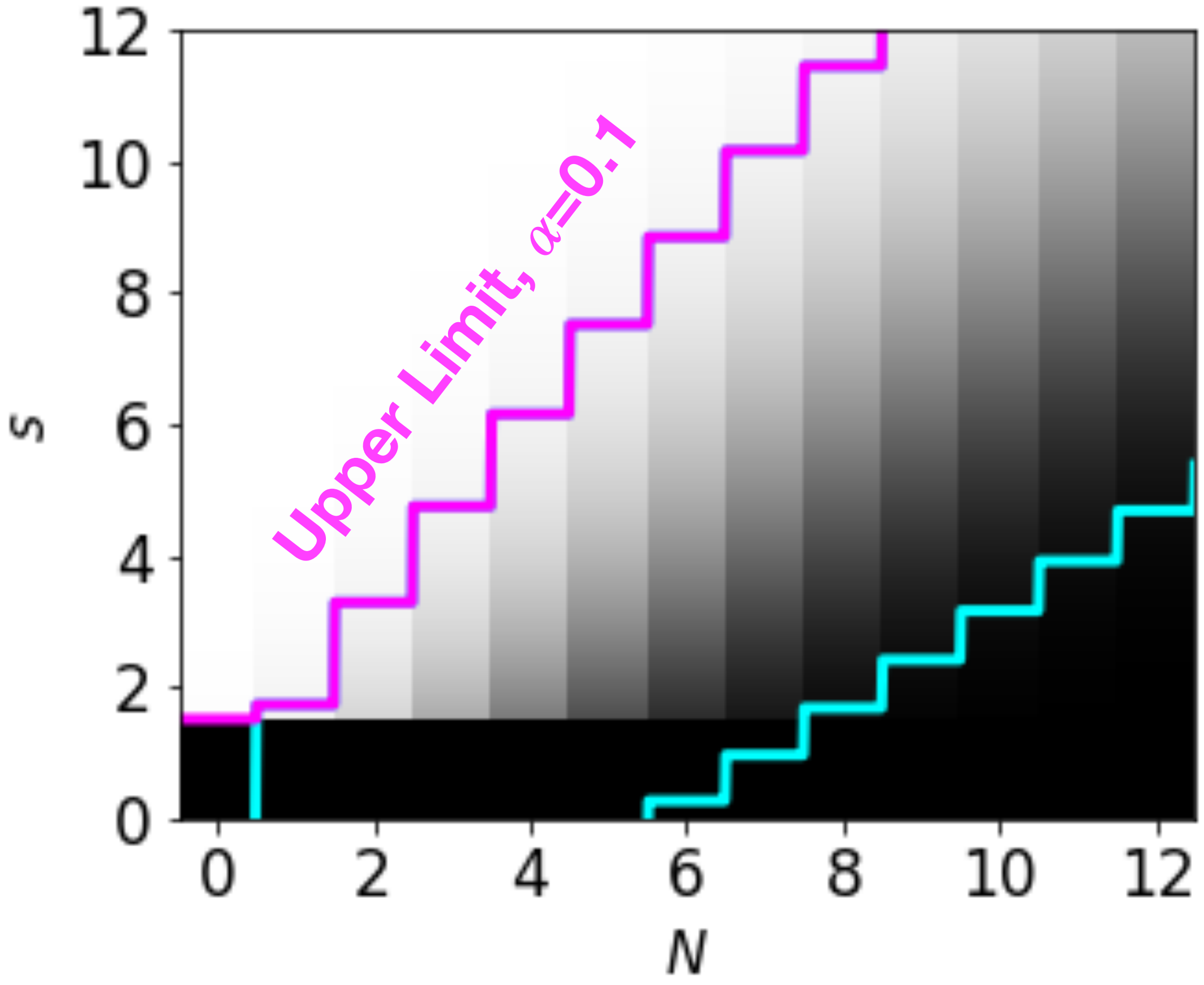
Knut Dundas Morå

CLs and PCL:

CLs



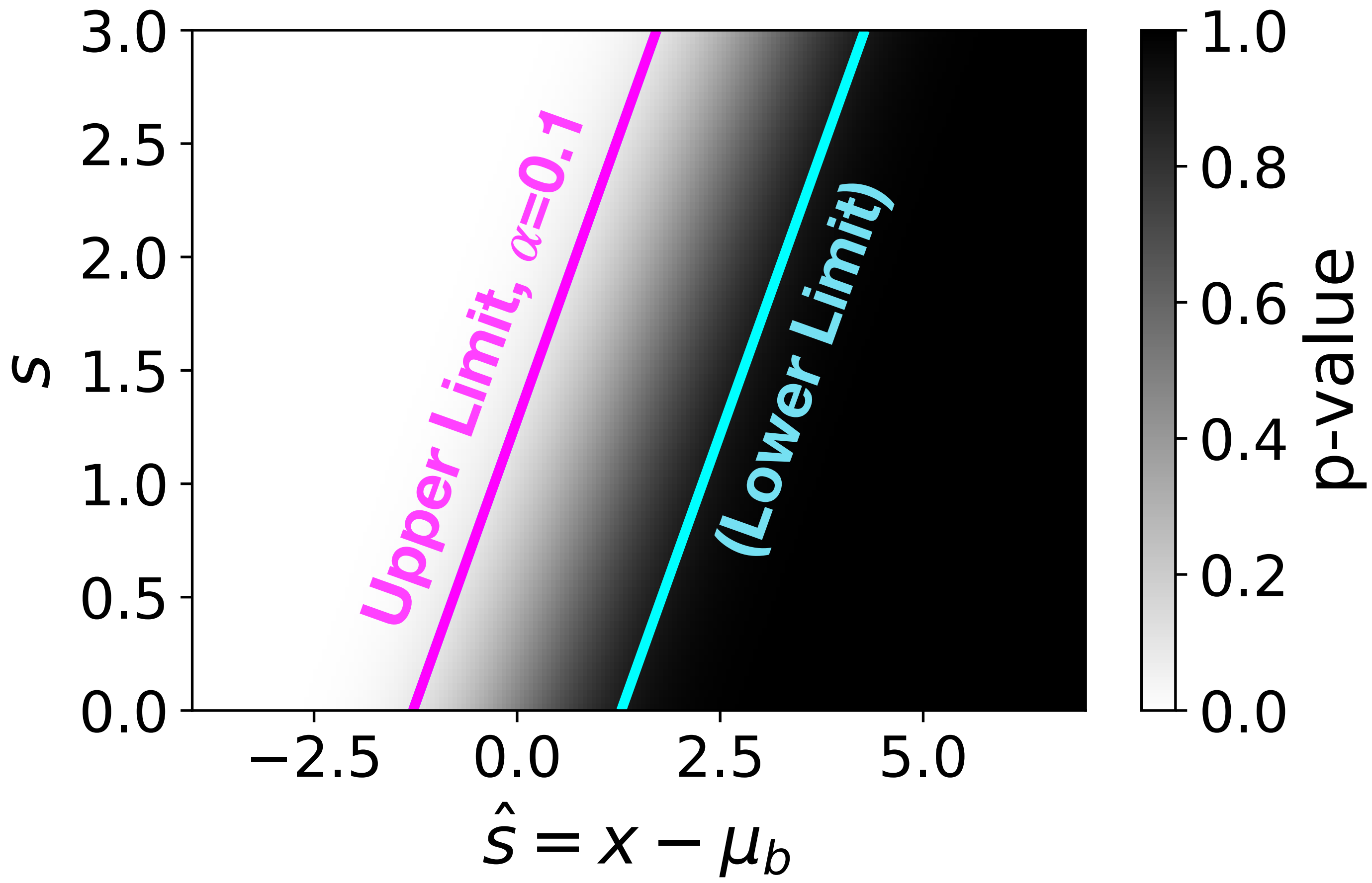
PCL



An upper-limit only construction must provide arbitrarily low upper limits

- A frequentist 1-α confidence interval contains the true value 1-α of repeated experiments

- An upper limit construction on a signal rate must exclude no signal in α of experiments

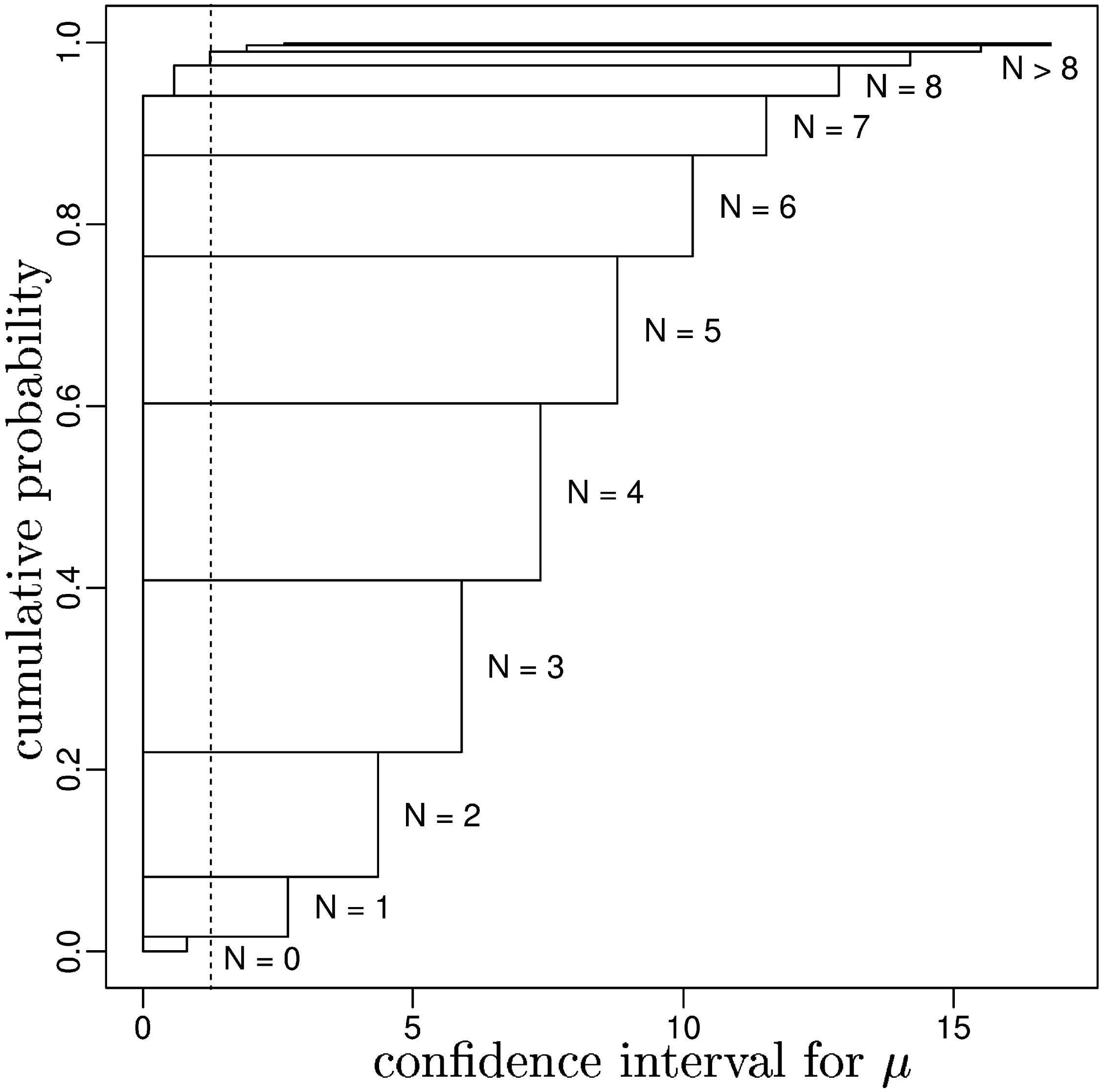


Neyman construction for a signal s
x ~ Gaussian with $\mu = s+\mu_b$ σ=1,

- Very small (or empty) intervals correspond to a large downwards fluctuation, which could be bad luck, or a mis-modelled background.

- An important point is that these intervals are relatively improbable (disregarding mis-modelling)

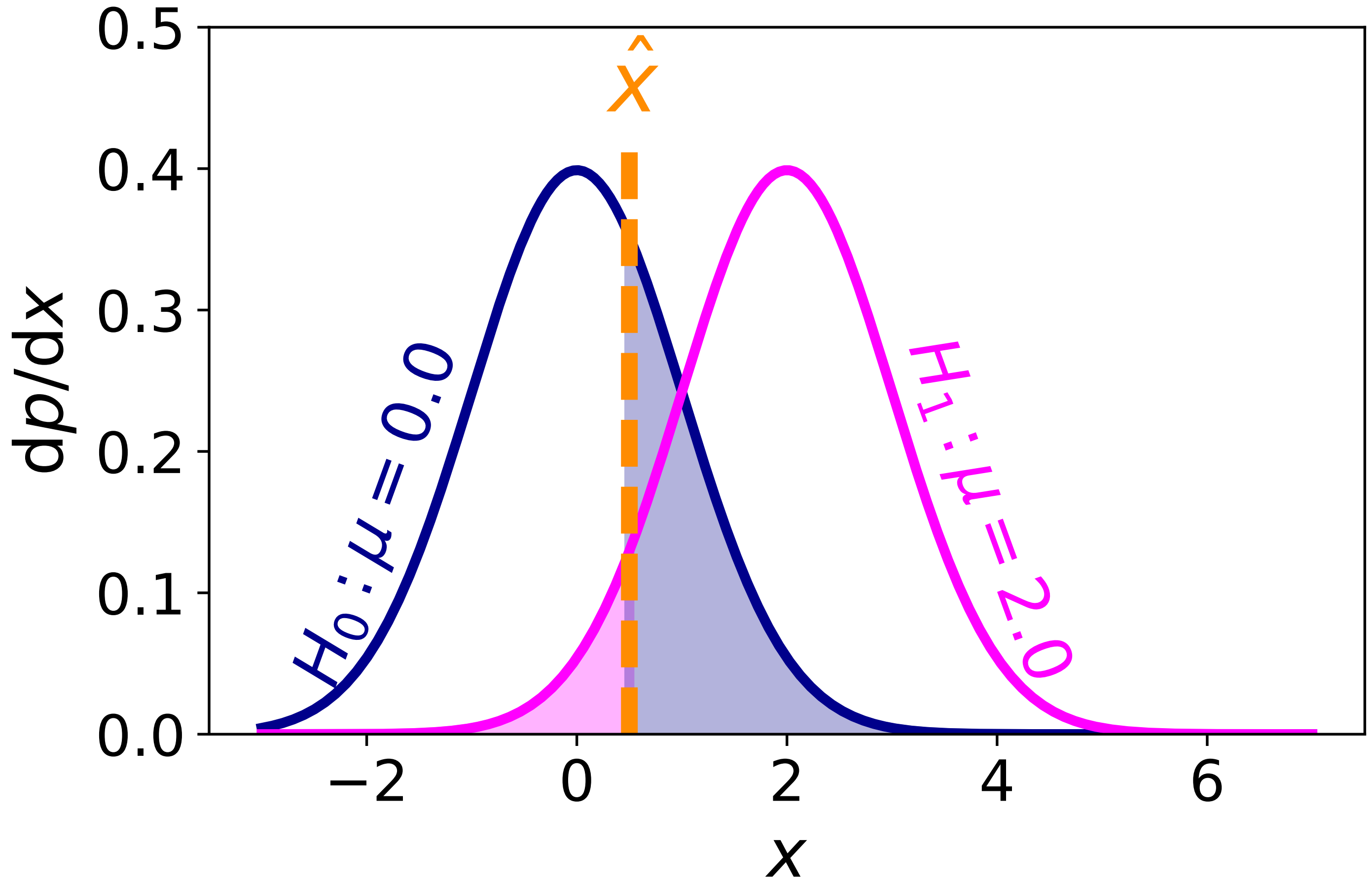
- Still, experiments have been loath to accept this risk (which, with perfect coverage is ~ α)
  - An unexplored solution is to adjust α until the risk is acceptable



From D. van Dyks comment to M. Mandelkern "Setting Confidence Intervals for Bounded Parameters" Stat. Science (2002)

The CLs approach

- The CLs method penalises the conventional p-value with increasing overlap between the test statistic distributions with and without signal.
  - For large significances, the result approaches the classical Neyman construction
- Typically constructed using the log-likelihood ratio as the test statistic
- Motivations originally included constructions of upper limits that agreed with Bayesian credible interval results using a flat prior
  - Also anticipated by the Helène formula that only applies to counting
- Contemporaneous alternatives included replacing the limit for under-fluctuations with the limit at x=0

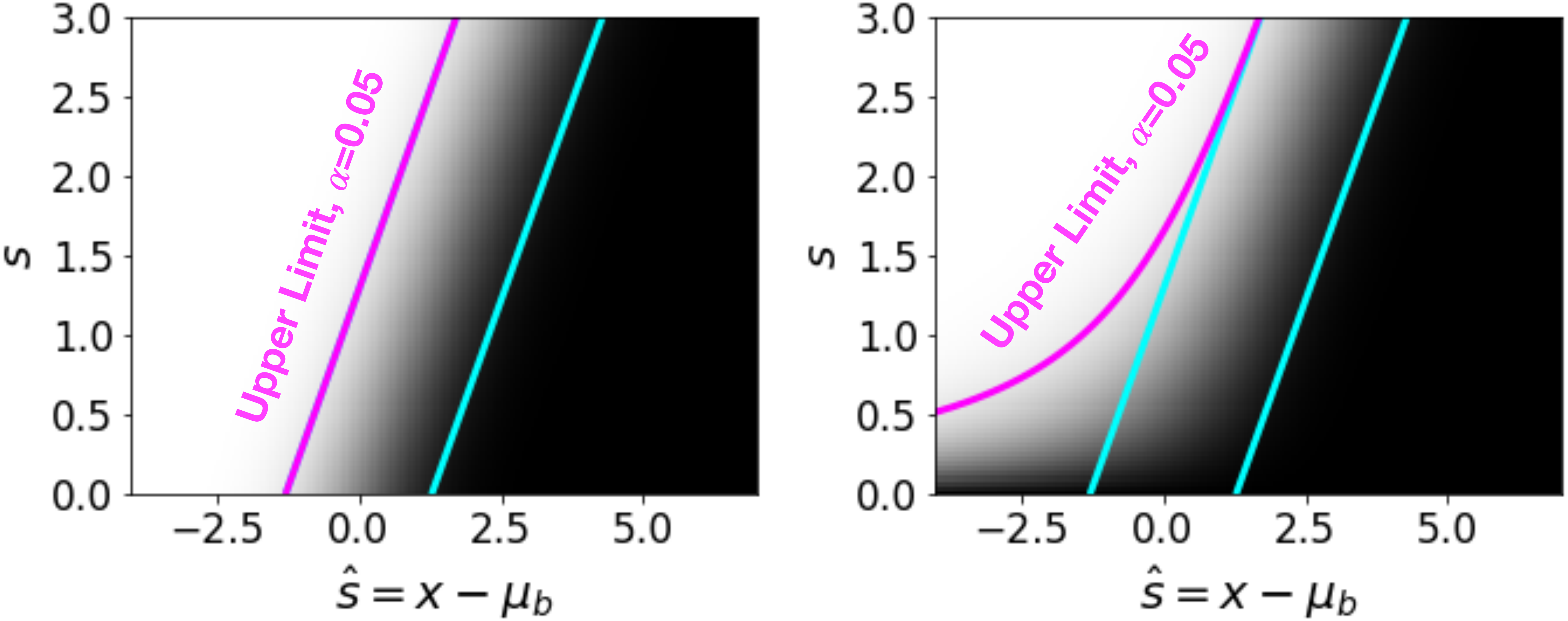


$$CL_{\rm s} \equiv \frac{p_{\rm s+b}}{1-p_{\rm b}} = \frac{P(x < \hat{x} | H_1)}{1 - P(\hat{x} > x | H_0)}$$

A. L. Read. Presentation of search results: the CLs technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, 2002.

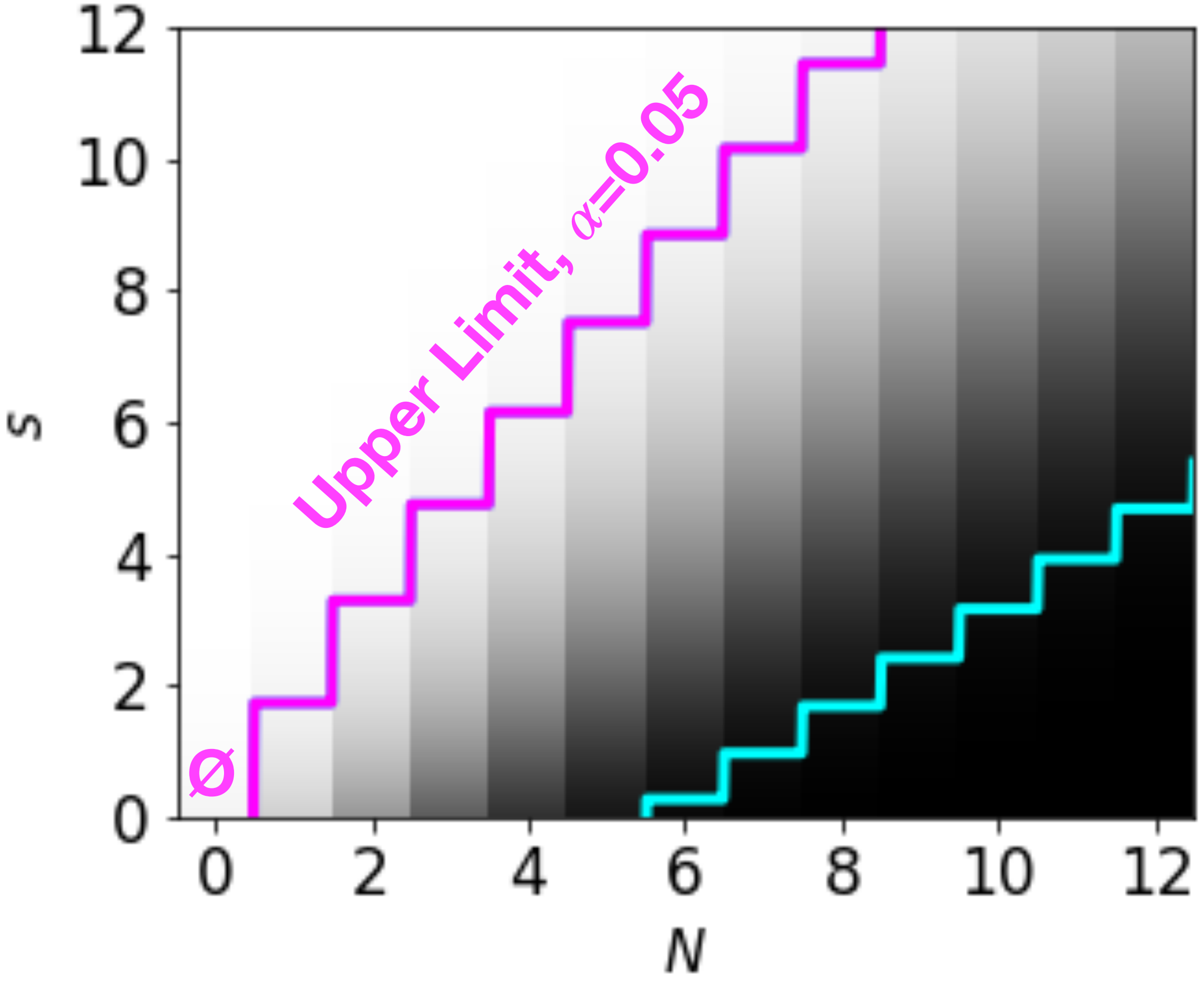
With a Gaussian:

Classical/Neyman **CLs**

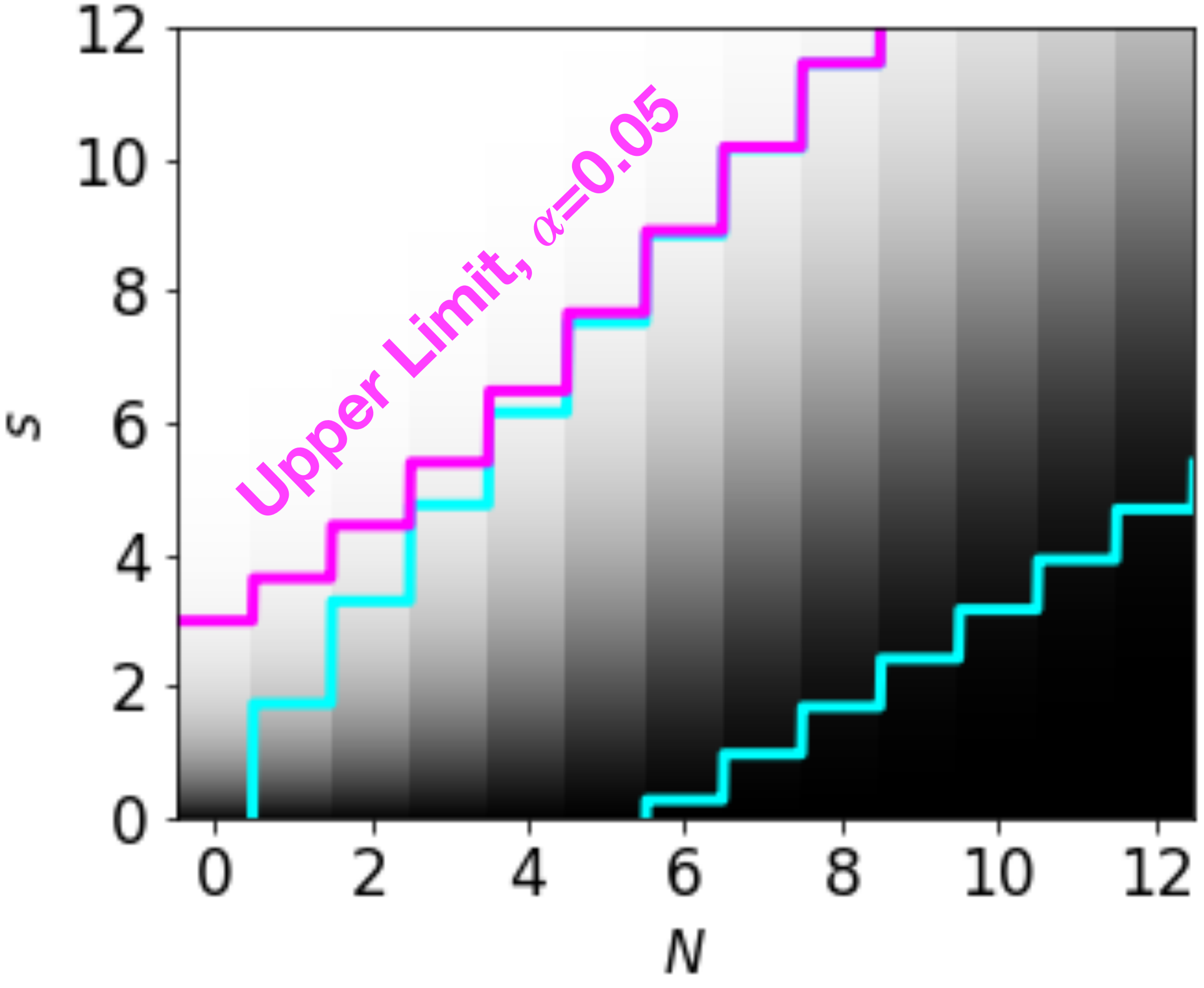


CLs for a counting experiment

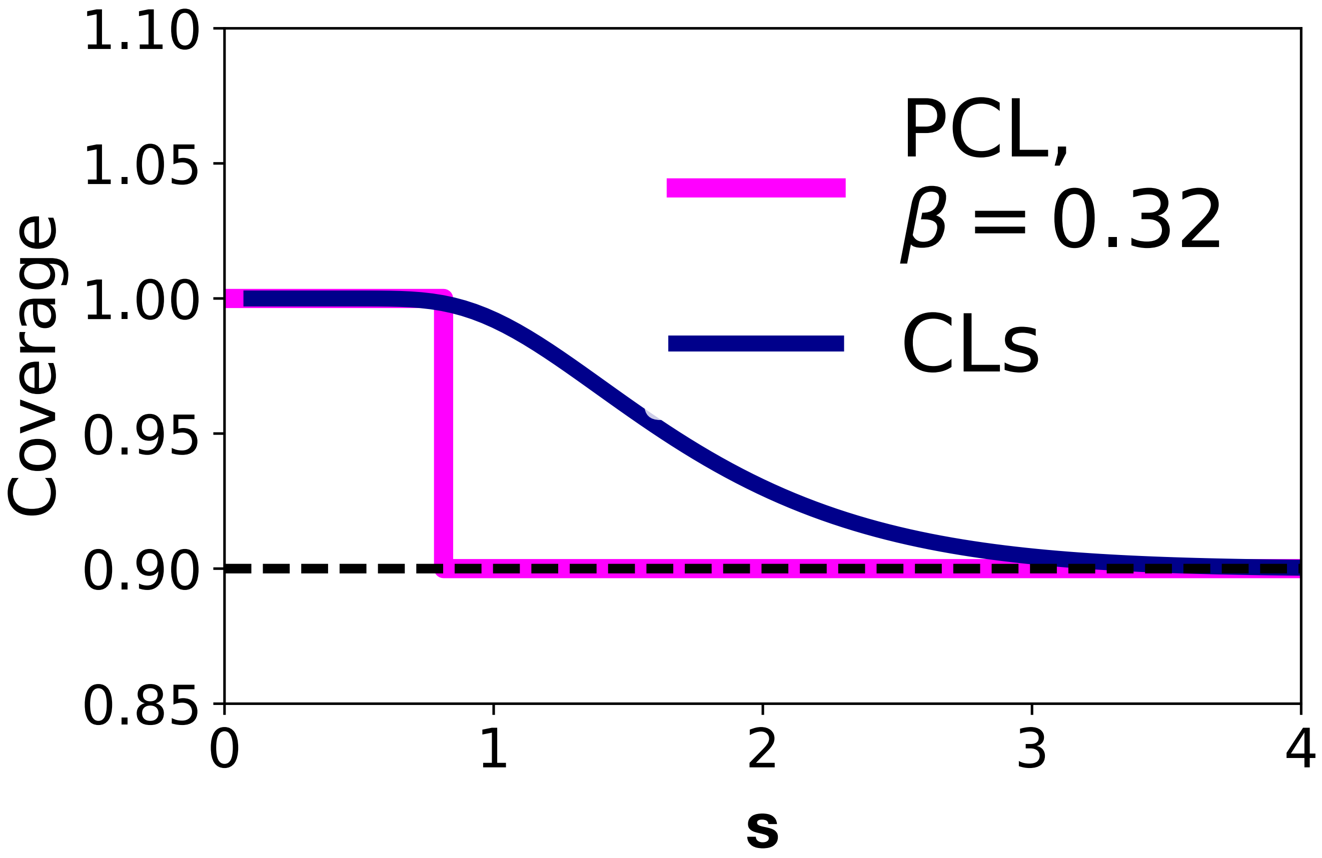
Classical/Neyman



CLs



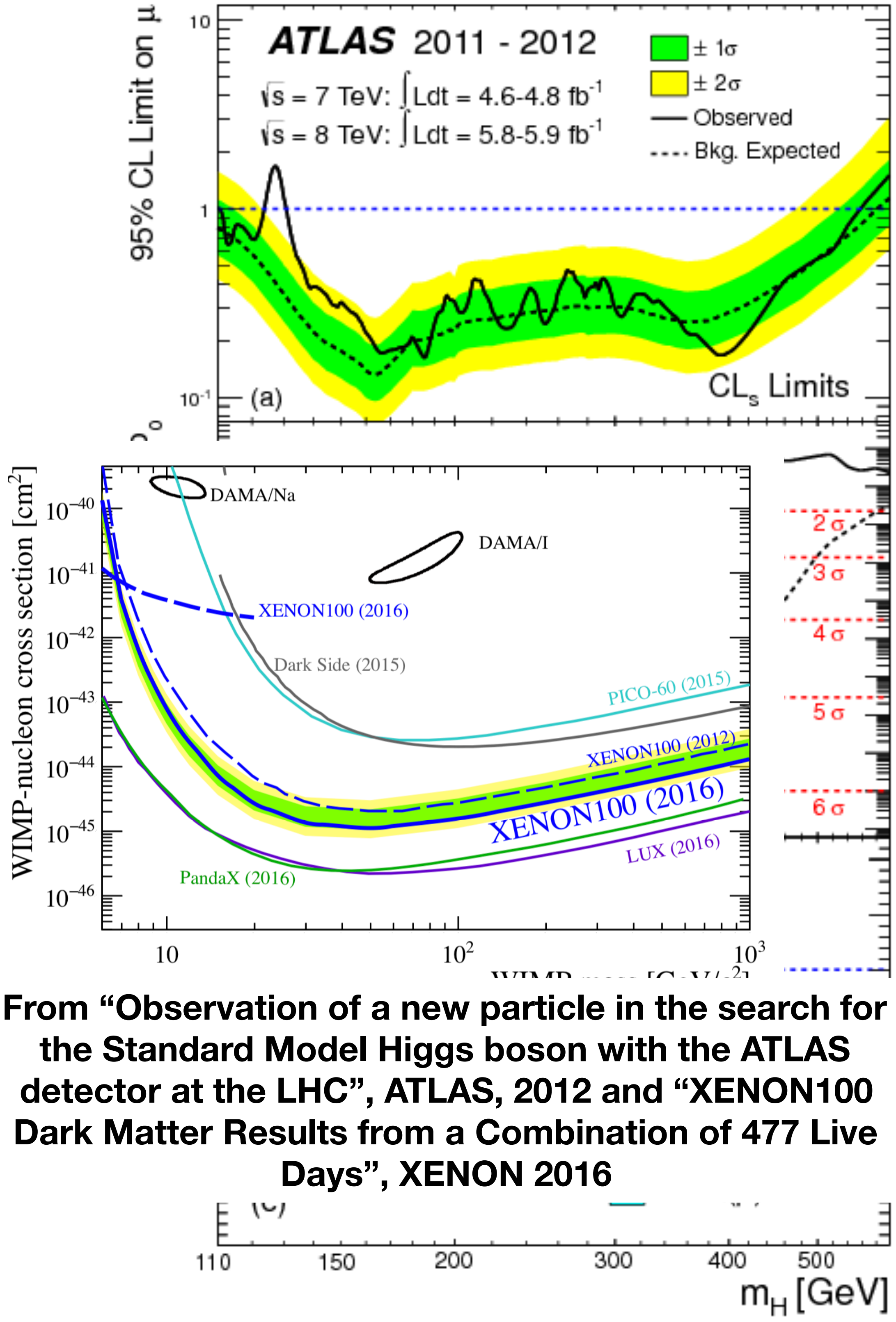
- The CLs overcoverage smoothly decreases from 1 when H1=H0 towards the nominal coverage
- The overcoverage / conservativeness extends above the median limit (1.63)



Coverage for the Gaussian example as function of signal expectation

Adoption

- CLs is widely adopted by LHC experiments
- Examples in direct detection include the XENON100 limit combination
  - Asymptotic results for the test statistic distribution were used to compute the CLs limits



From "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", ATLAS, 2012 and "XENON100 Dark Matter Results from a Combination of 477 Live Days", XENON 2016

G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: Eur. Phys. J.C73,2501(2013)].

Power-constrained limits

- The PCL approach is to require the experiment to have a minimum discovery power $\boldsymbol{\beta}$ for each model that it excludes

- If the un-constrained confidence interval construction yields a lower upper limit, that limit is truncated at the signal strength with the minimal discovery power

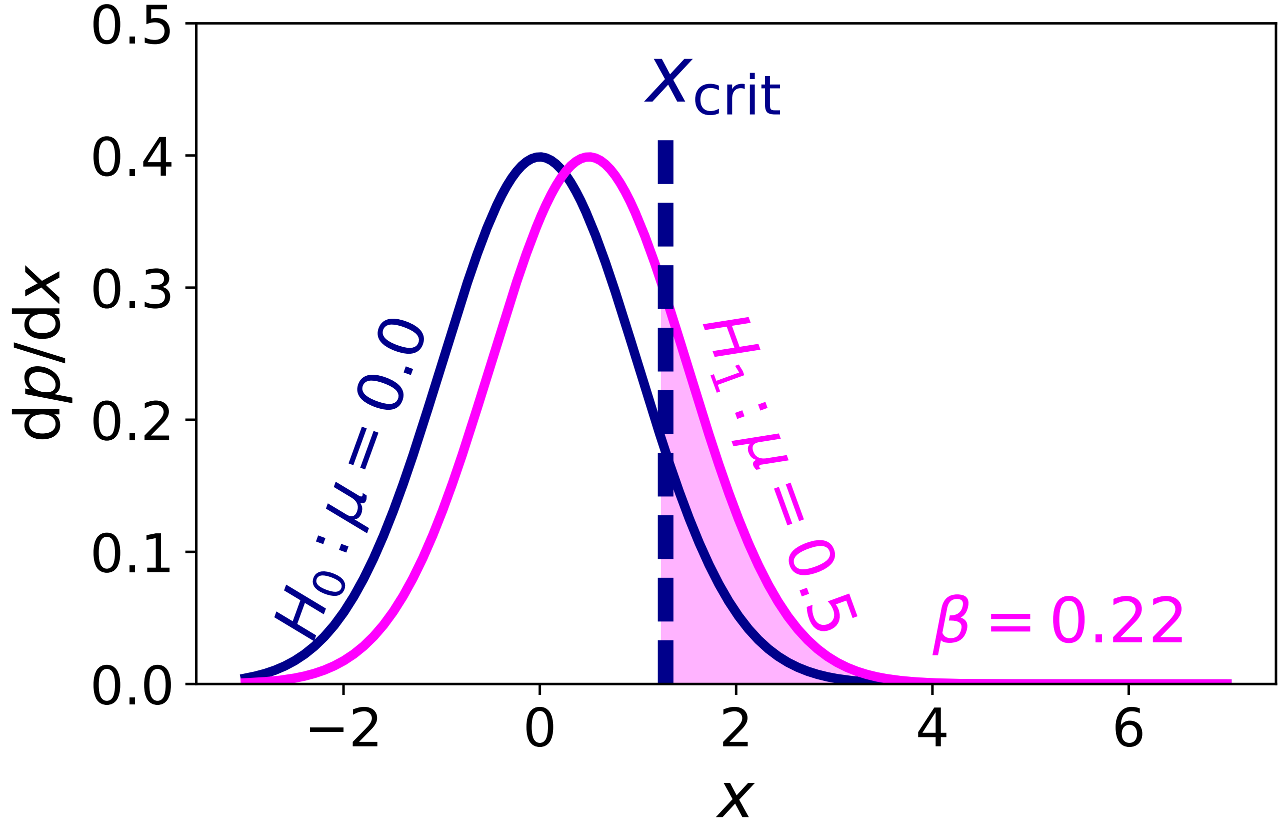


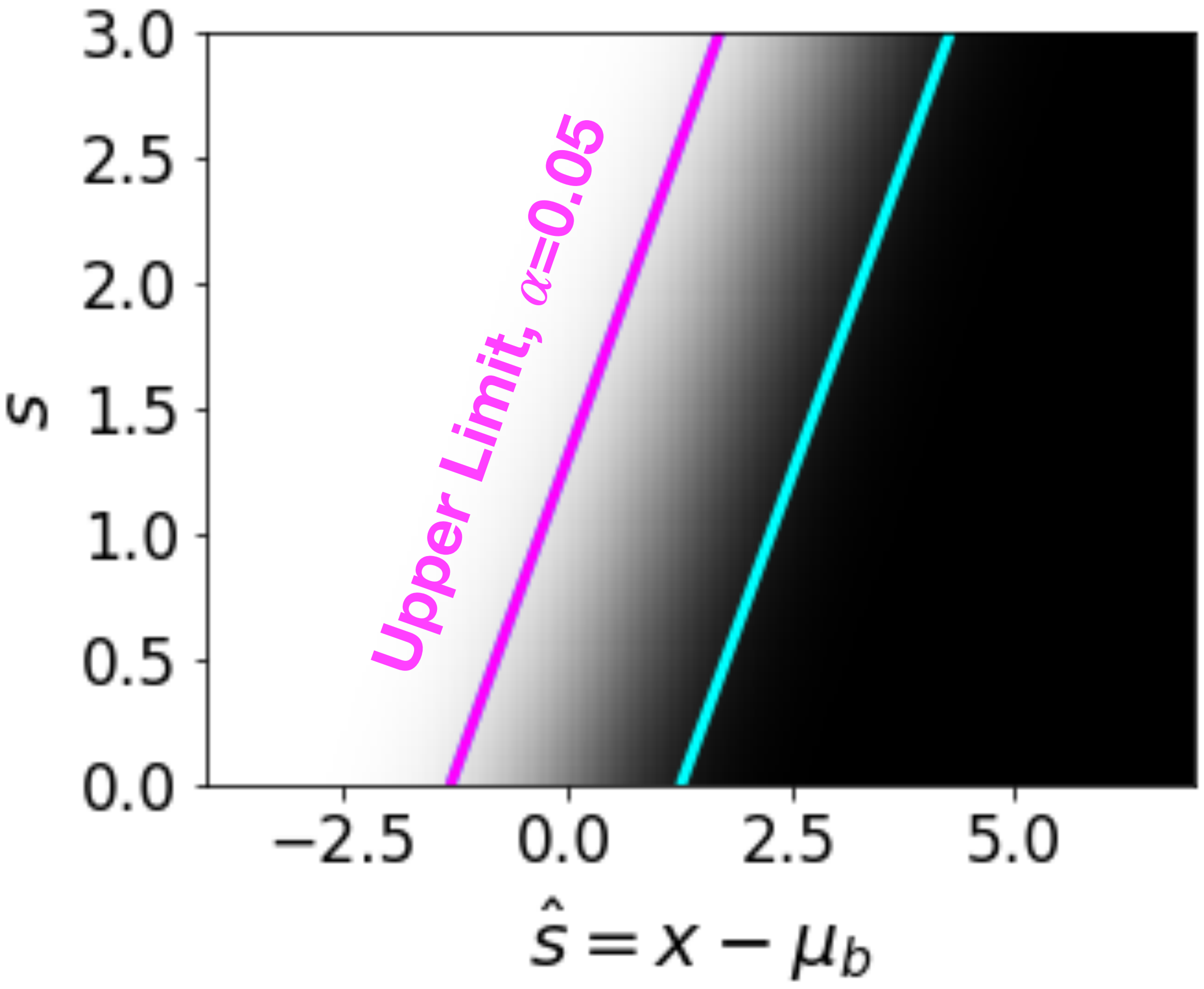
Illustration of discovery power

G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Power-Constrained Limits. *pre-print*, 2011. [physics.data-an/1105.3166].

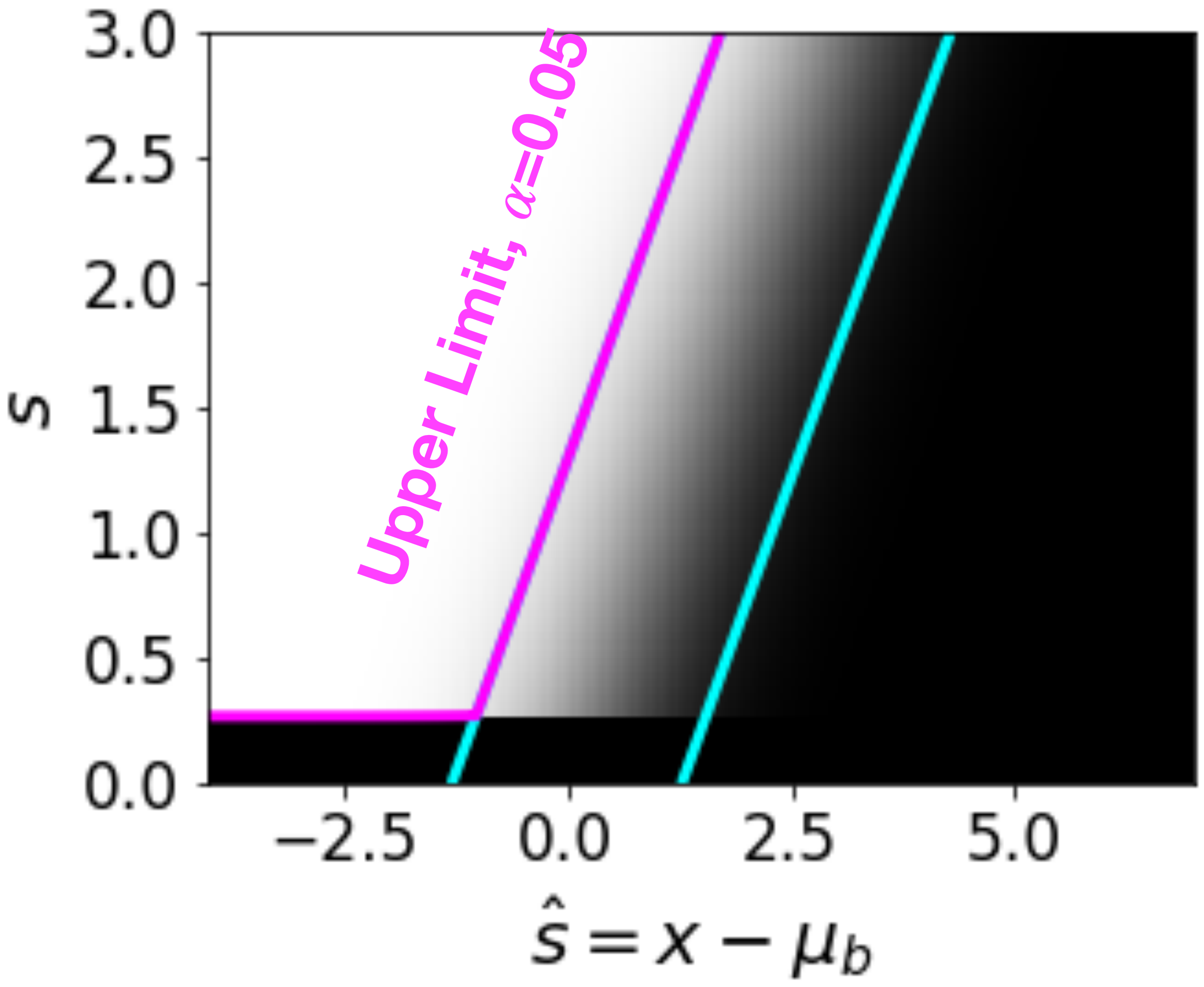
G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Power-Constrained Limits. *pre-print*, 2011. [physics.data-an/1105.3166].

With a Gaussian:

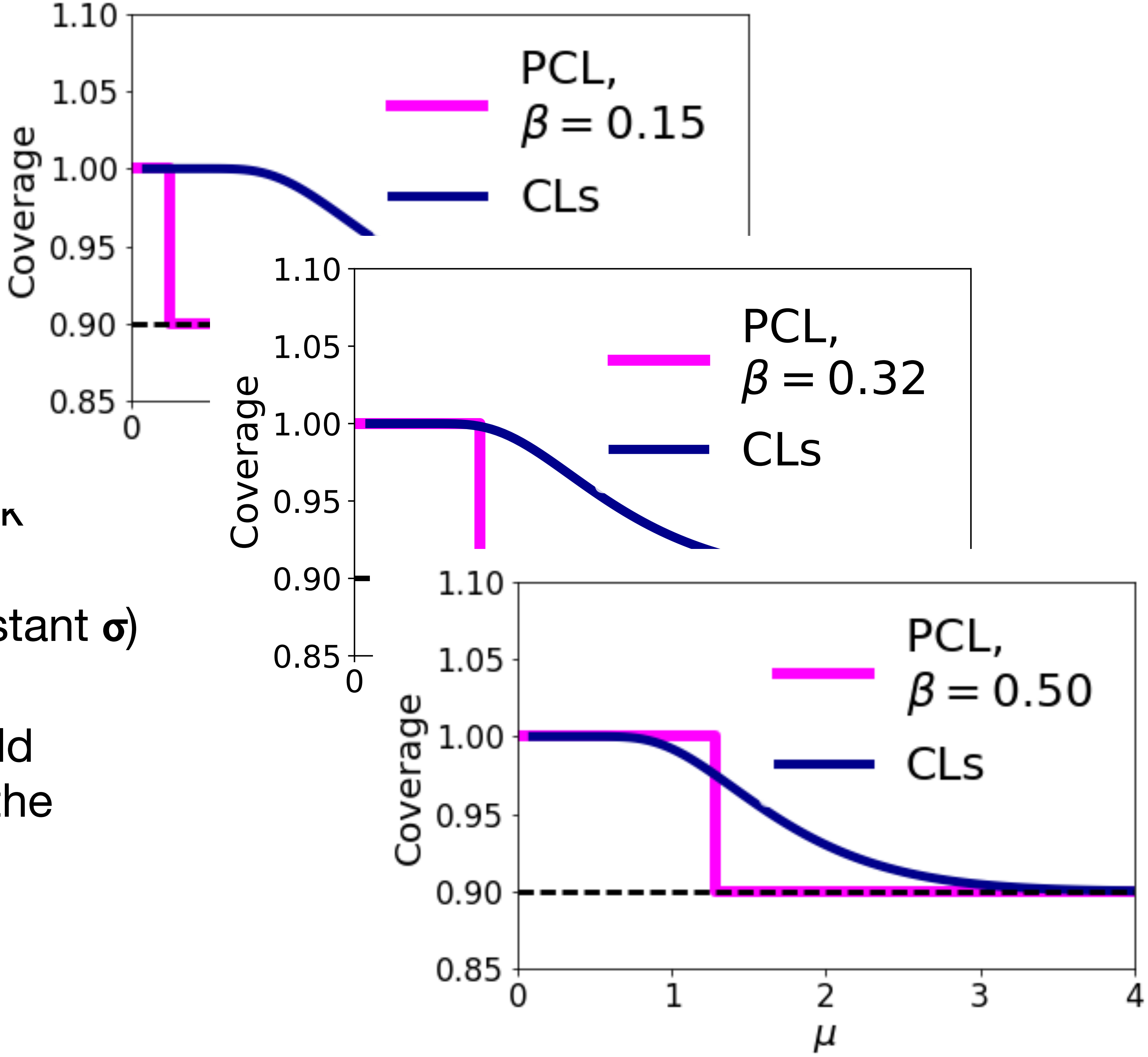
Classical/Neyman



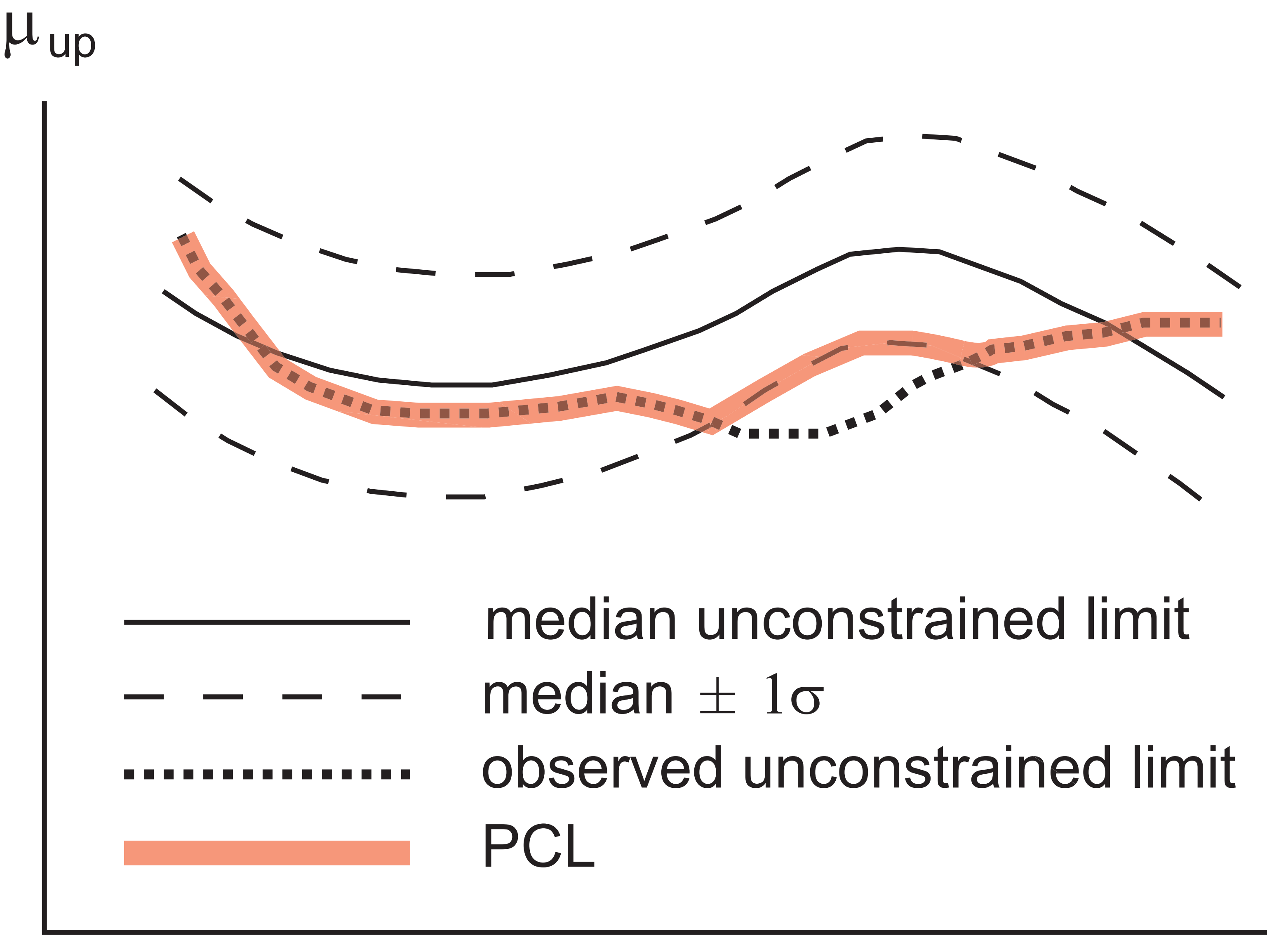
PCL



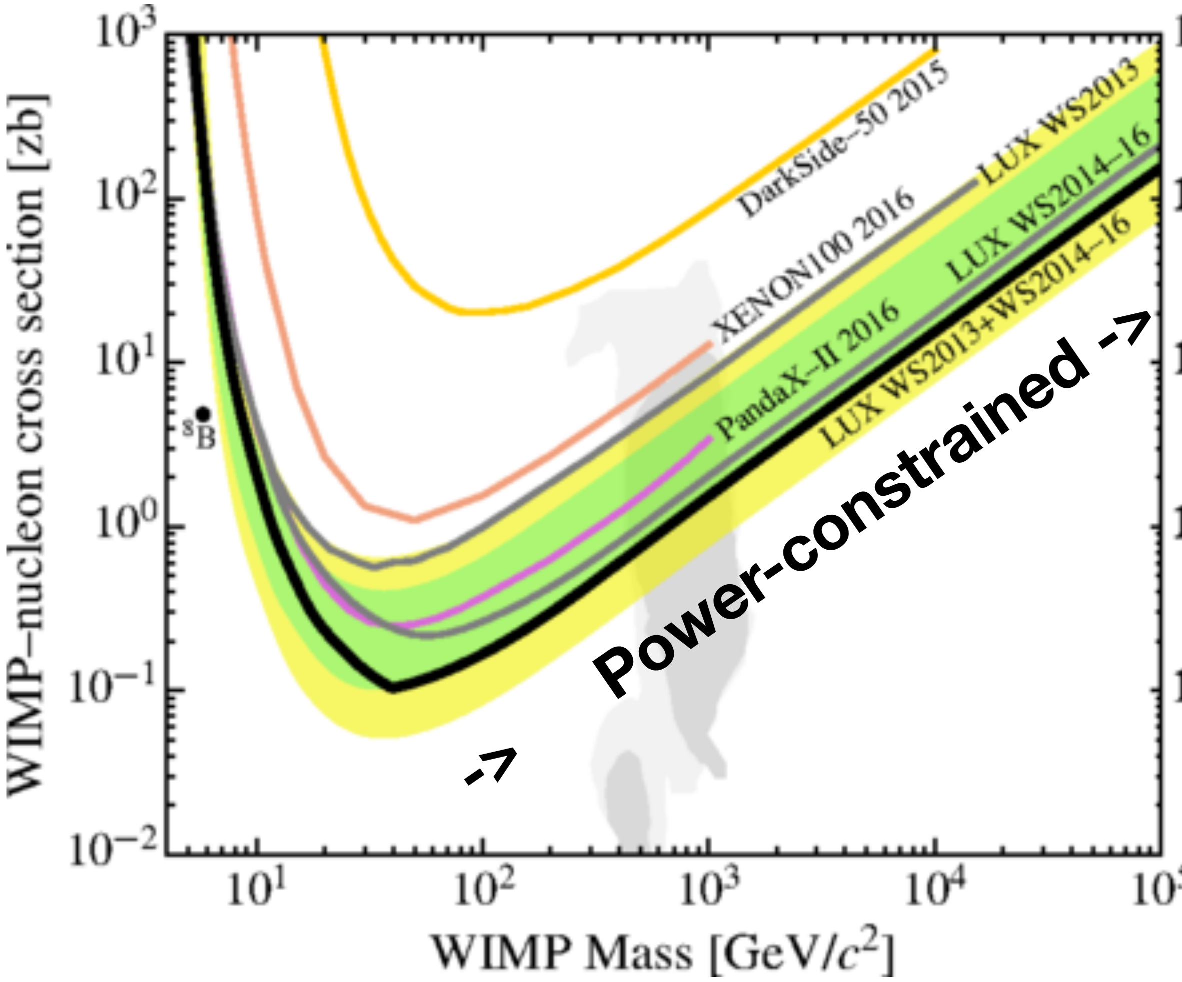
- The threshold discovery
  a free parameter of the m
  - May be considered bot
    benefit and a draw-back
  - For a Gaussian (of constant **σ**) measurement or limit distribution the threshold power corresponds to the percentile of the limit distribution
  - Suggested: 0.158— corresponding to a 1-sigma downwards fluctuation



Serving suggestion and usage



From *Power-Constrained Limits* by Cowan, Cranmer, Gross & Vitells



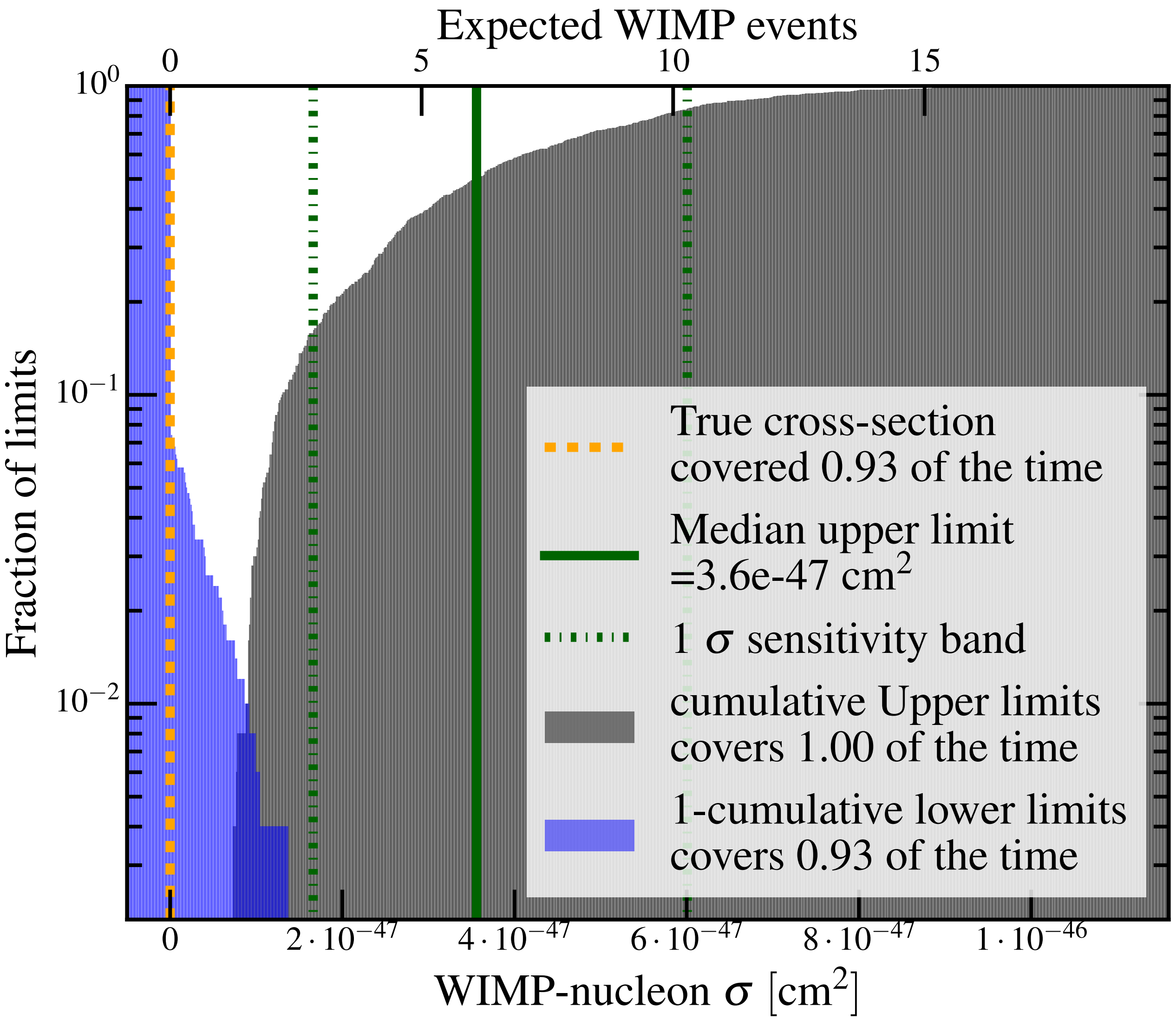
Power-constrained ->

->

Results from a Search for Dark Matter in the Complete Lux Exposure, PRL 118 (2017)

Two-sided intervals

- Unified confidence intervals give both one- or two-sided intervals, based on the ordering parameter:
  $$R(\theta) = 2 \cdot \log\left[\mathcal{L}(\hat{s})/\mathcal{L}(s)\right]$$
- For these constructions, the coverage for small signals is kept since the confidence interval will exclude s=0 for p-values below α
- LUX, PandaX-II and XENON all use unified interval constructions, as well as applying a PCL

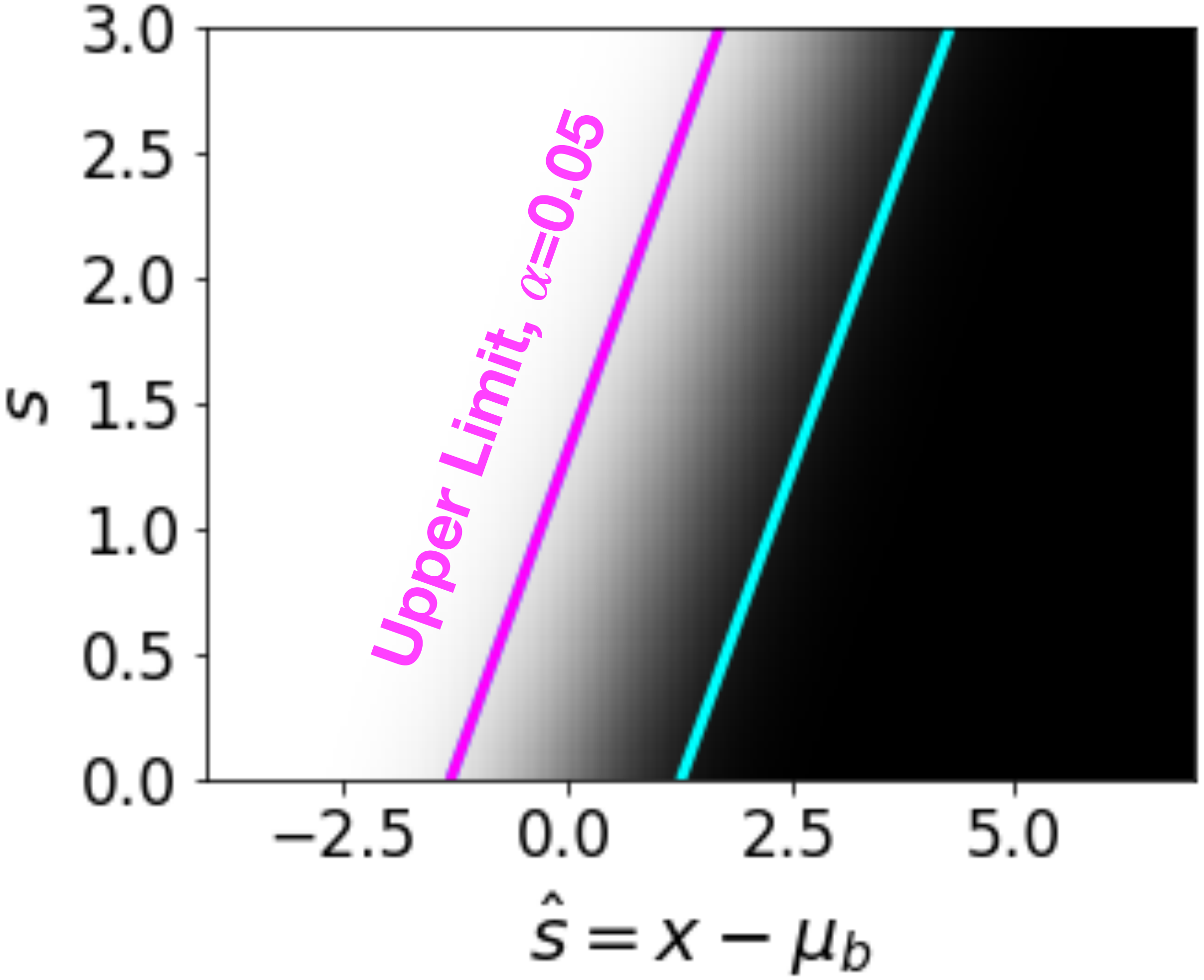


Distribution of upper and lower limits for the XENON1T 1tonne-year SI search

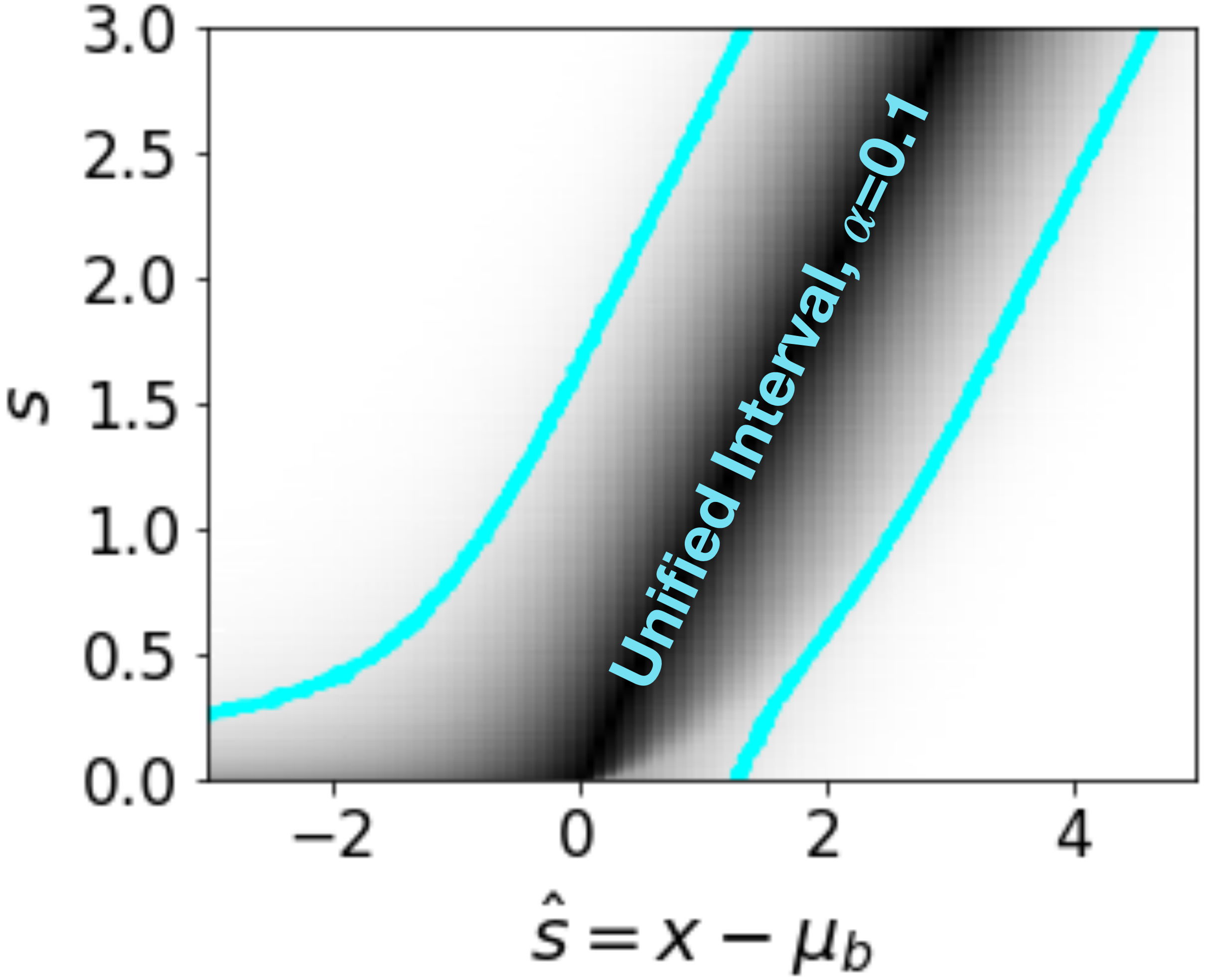
G. J. Feldman and R. D. Cousins. A Unified approach to the classical statistical analysis of small signals. *Phys. Rev.*, D57:3873–3889, 1998.

With a Gaussian:

Classical/Neyman



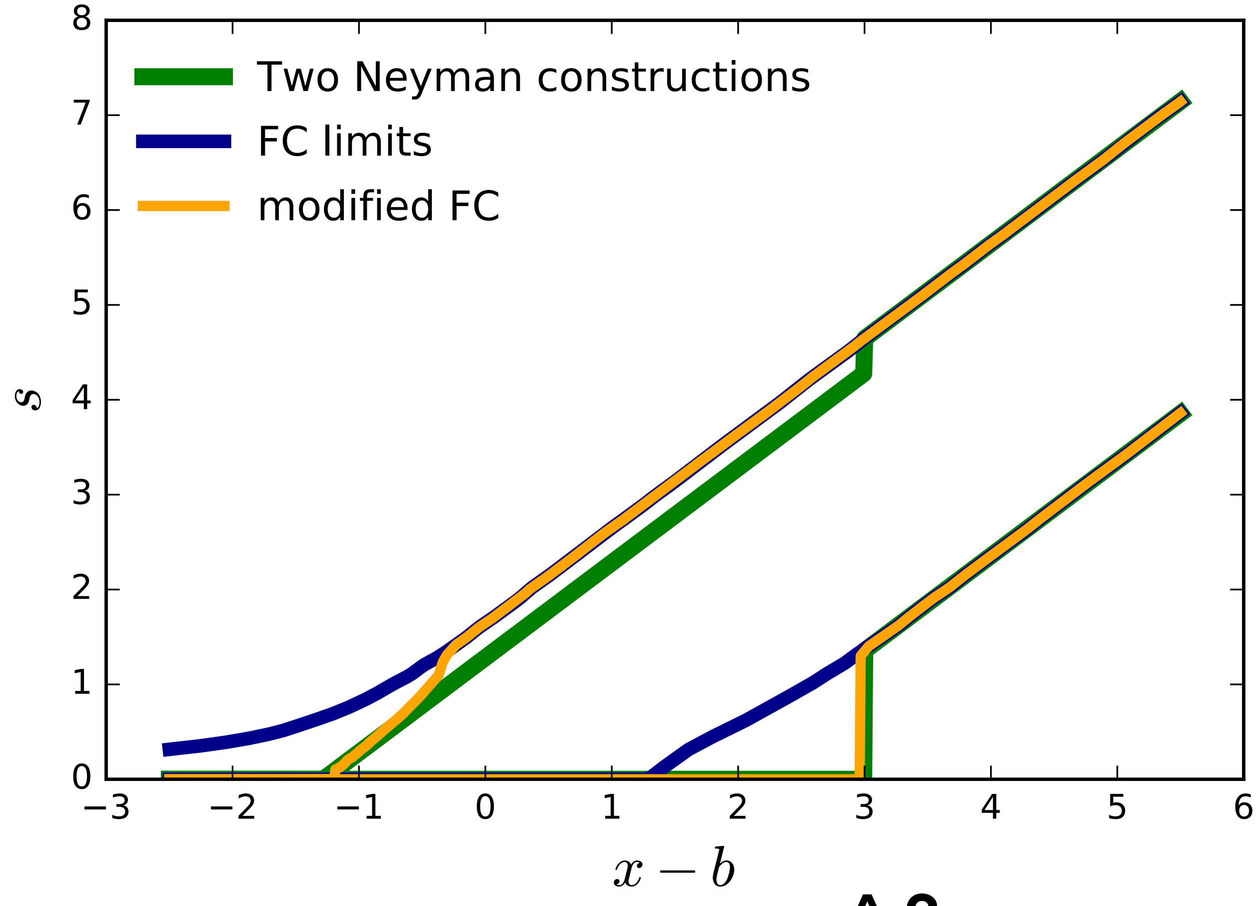
Unified Interval



Discovery thresholds

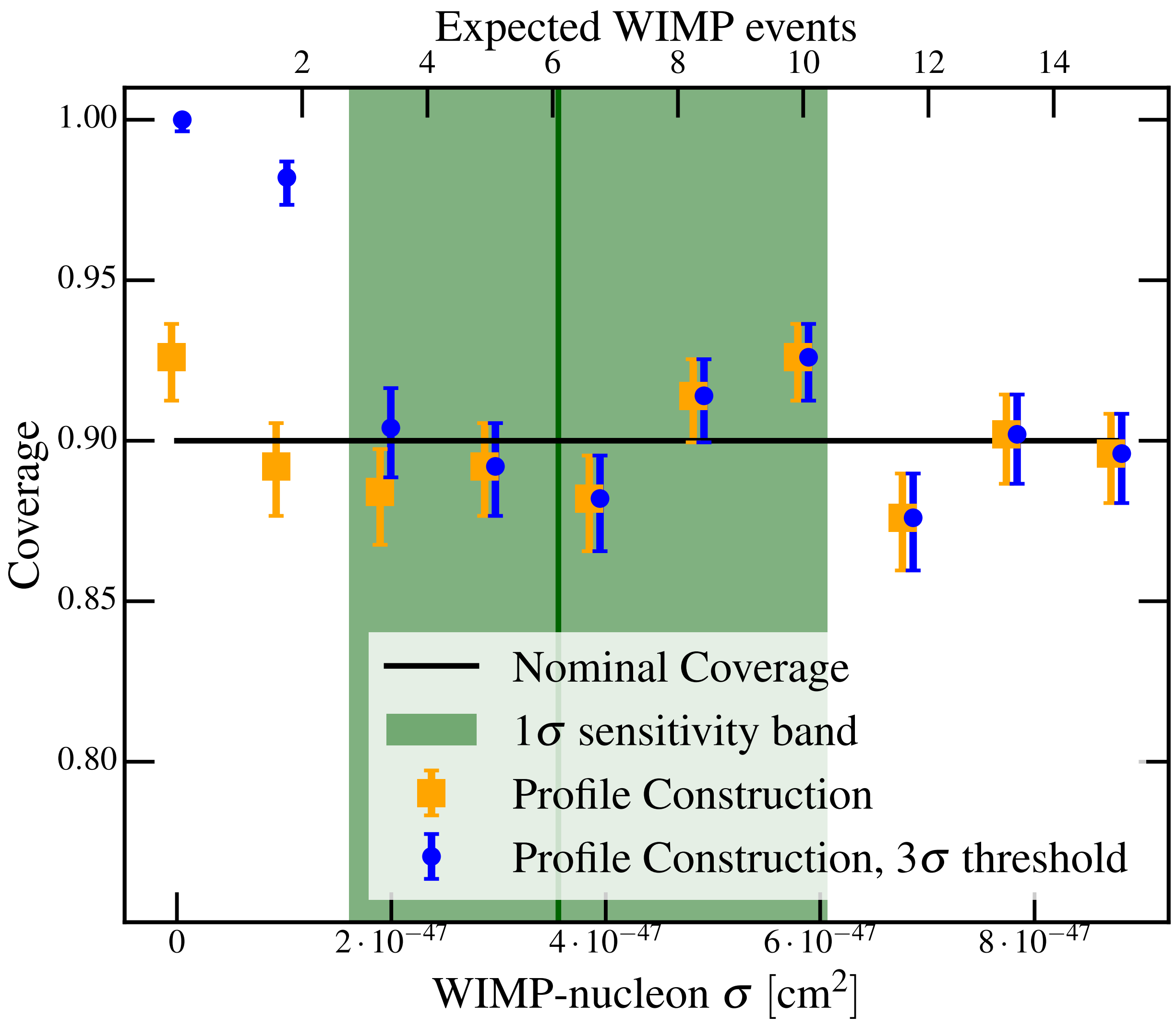
- For the 1 tonne-year XENON1T results, the collaboration resolved to only report the upper edge of the confidence interval until the discovery significance exceeds 3 **σ**

- It is possible to modify the unified interval to reduce the overcoverage due to this, at the cost of requiring PCL or a similar solution for low signals



^ 3 σ threshold

- The discovery significance threshold gives overcoverage for signal sizes ~below the sensitivity band



ToyMC-computation of XENON1T SI coverage for a 50GeV WIMP

- CLs is widely adopted in particle physics, and does not require additional fiducial choices

- PCL emphasises the coverage properties of the construction, and (for most choices of power threshold) imposes less over-coverage than the CLs method

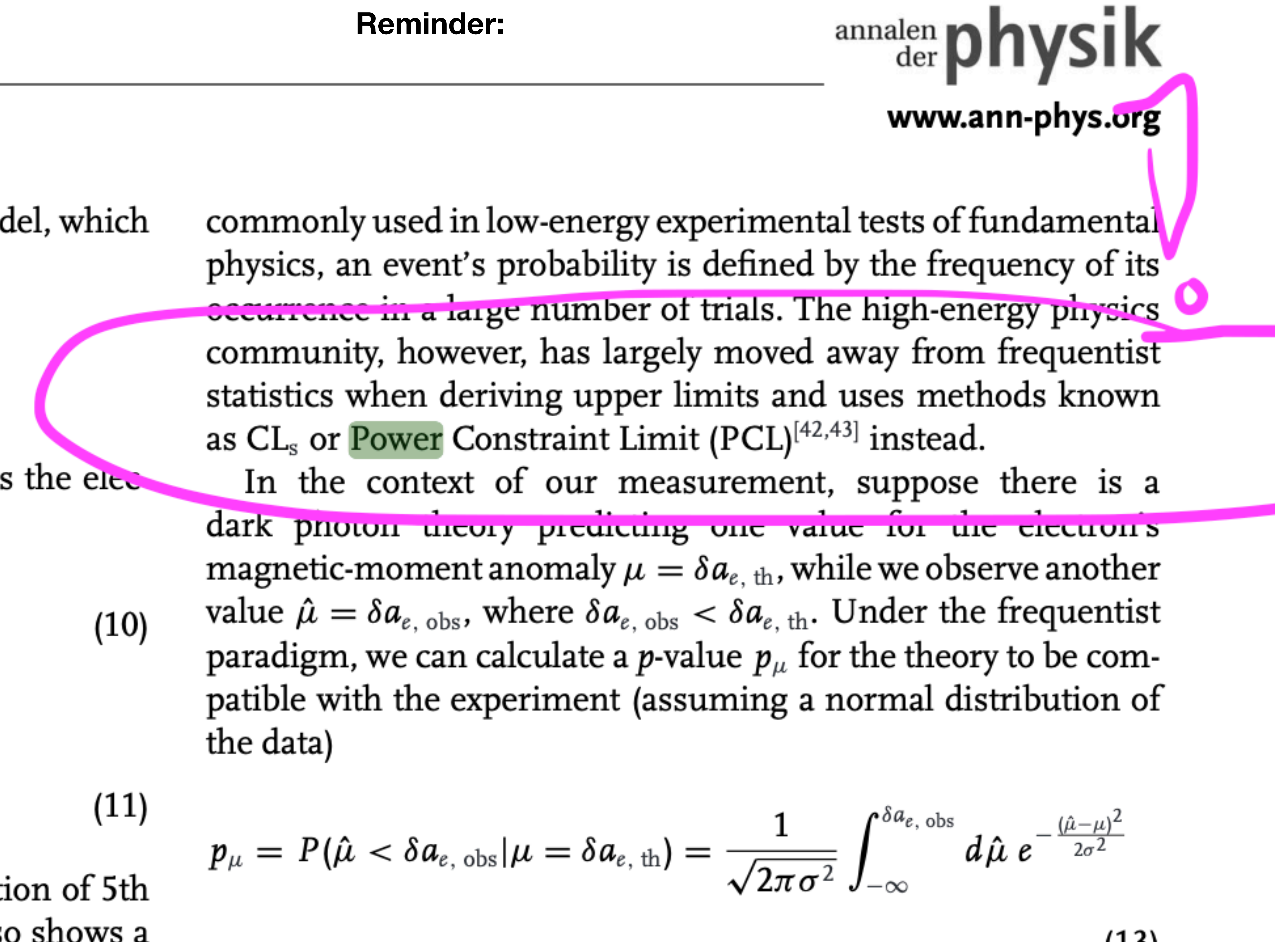
- Unified confidence intervals do not return arbitrarily low limits, but liquid xenon TPC collaborations have still combined them with PCL to avoid (the small probability of) limits of ~1 event.

- Regardless of the choice, additional norms or rules-of-thumb for goodness-of-fit would be of great help.

Summary

- Limit-only 1- α confidence level constructions will exclude even the no-signal case α of the time.
- Direct detection experiments have accepted over-coverage for signals for which they have little sensitivity;
  - By applying the CLs method of modifying p-values
  - By setting a threshold, based on discovery power, below which they will not set lower limits (PCL)
- The unified interval construction achieves coverage for even vanishing signals by having $\boldsymbol{\alpha}$ of intervals be two-sided.
  - To avoid even this, XENON1T set a discovery significance threshold (3 σ rather than p< $\boldsymbol{\alpha}$) for reporting two-sided confidence intervals.

Reminder:



!

del, which

s the elec

(10)

(11)

ion of 5th

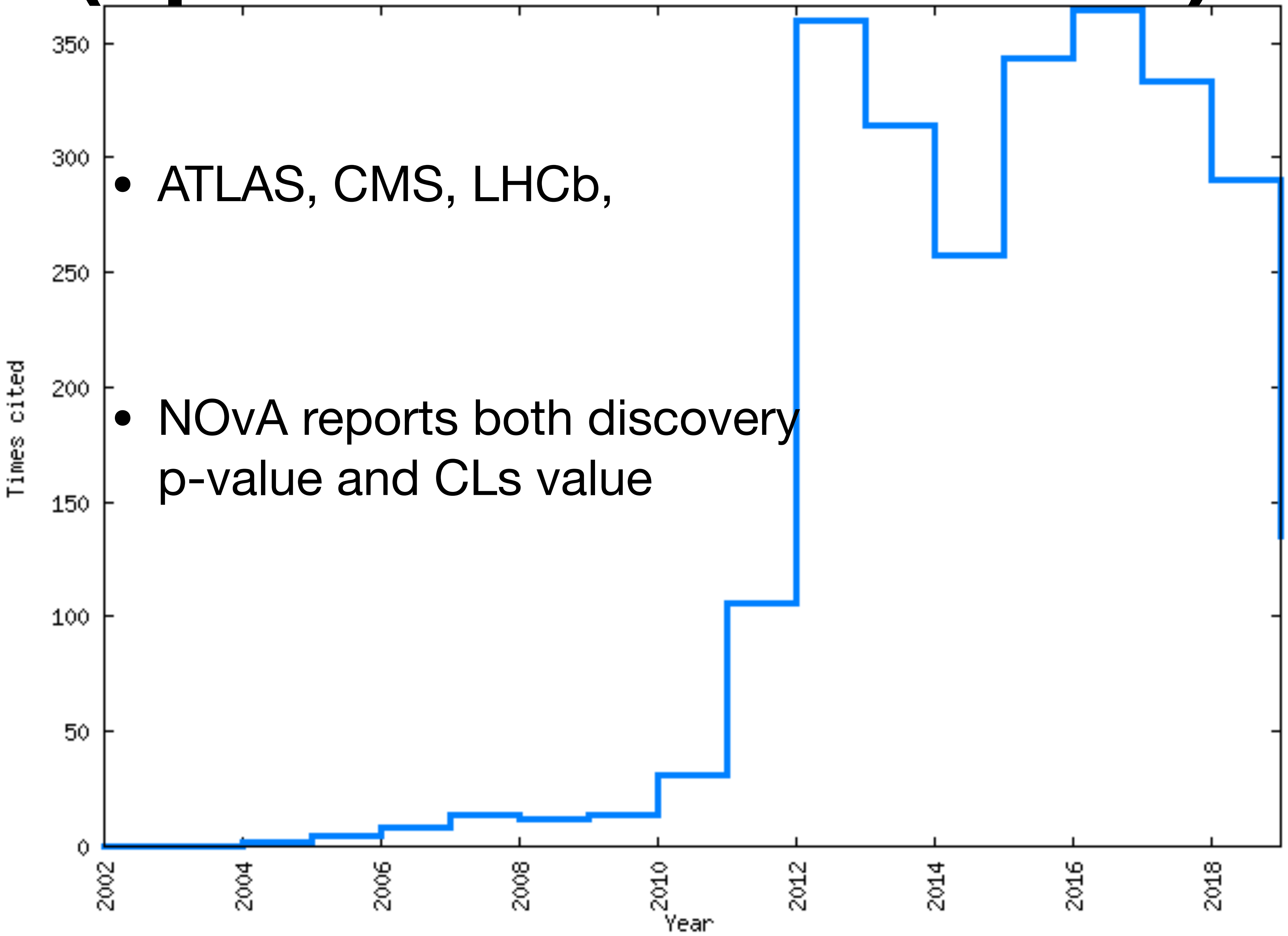
o shows a

commonly used in low-energy experimental tests of fundamental physics, an event's probability is defined by the frequency of its occurrence in a large number of trials. The high-energy physics community, however, has largely moved away from frequentist statistics when deriving upper limits and uses methods known as $\mathrm{CL_s}$ or Power Constraint Limit (PCL)[42,43] instead.

In the context of our measurement, suppose there is a dark photon theory predicting one value for the electron's magnetic-moment anomaly $\mu = \delta a_{e,\,\mathrm{th}}$, while we observe another value $\hat{\mu} = \delta a_{e,\,\mathrm{obs}}$, where $\delta a_{e,\,\mathrm{obs}} < \delta a_{e,\,\mathrm{th}}$. Under the frequentist paradigm, we can calculate a p-value p_μ for the theory to be compatible with the experiment (assuming a normal distribution of the data)

$$p_\mu = P(\hat{\mu} < \delta a_{e,\,\mathrm{obs}} | \mu = \delta a_{e,\,\mathrm{th}}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\delta a_{e,\,\mathrm{obs}}} d\hat{\mu}\; e^{-\frac{(\hat{\mu}-\mu)^2}{2\sigma^2}}$$

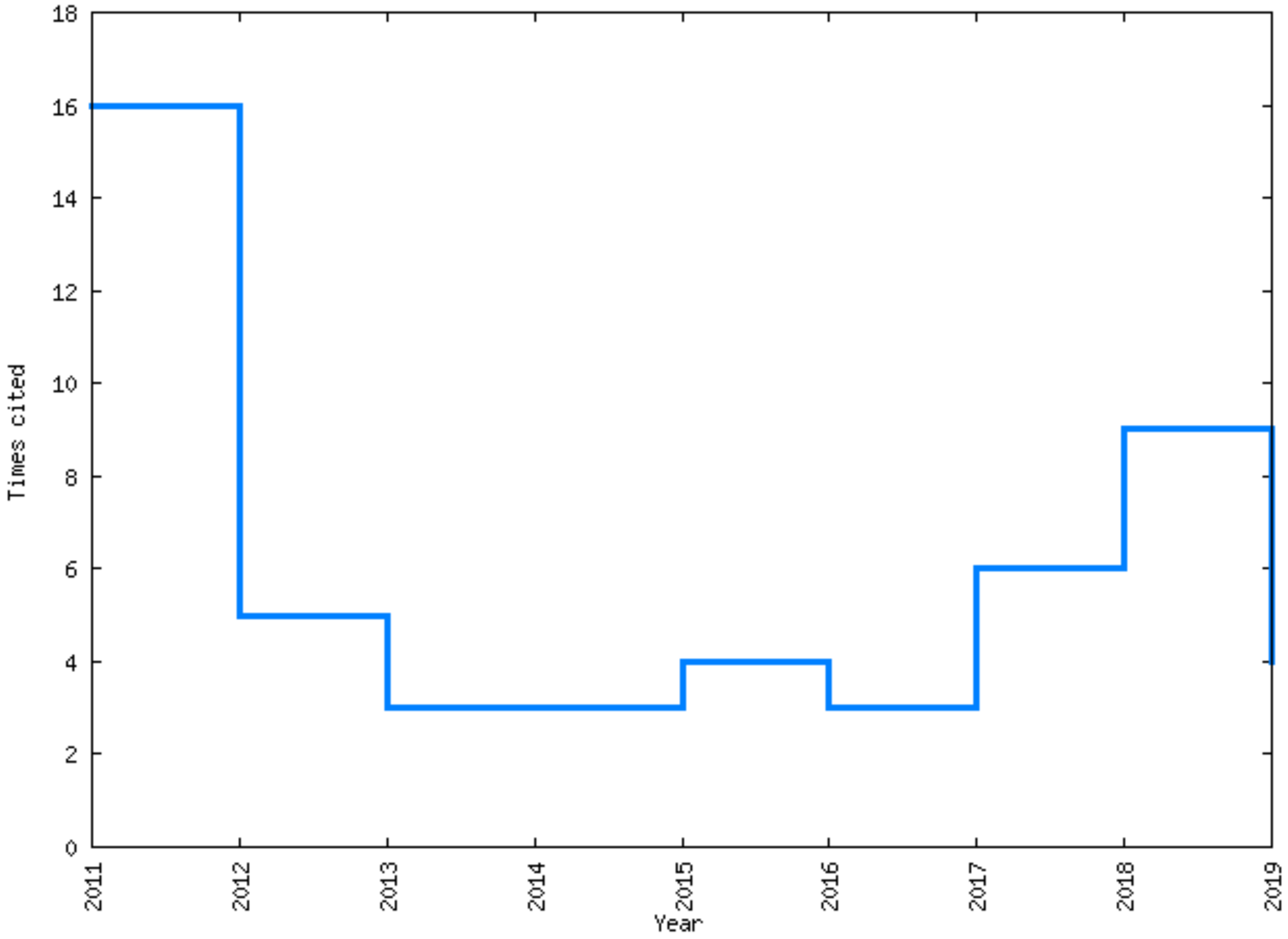
CLs citation history (spot the LHC run 1?)

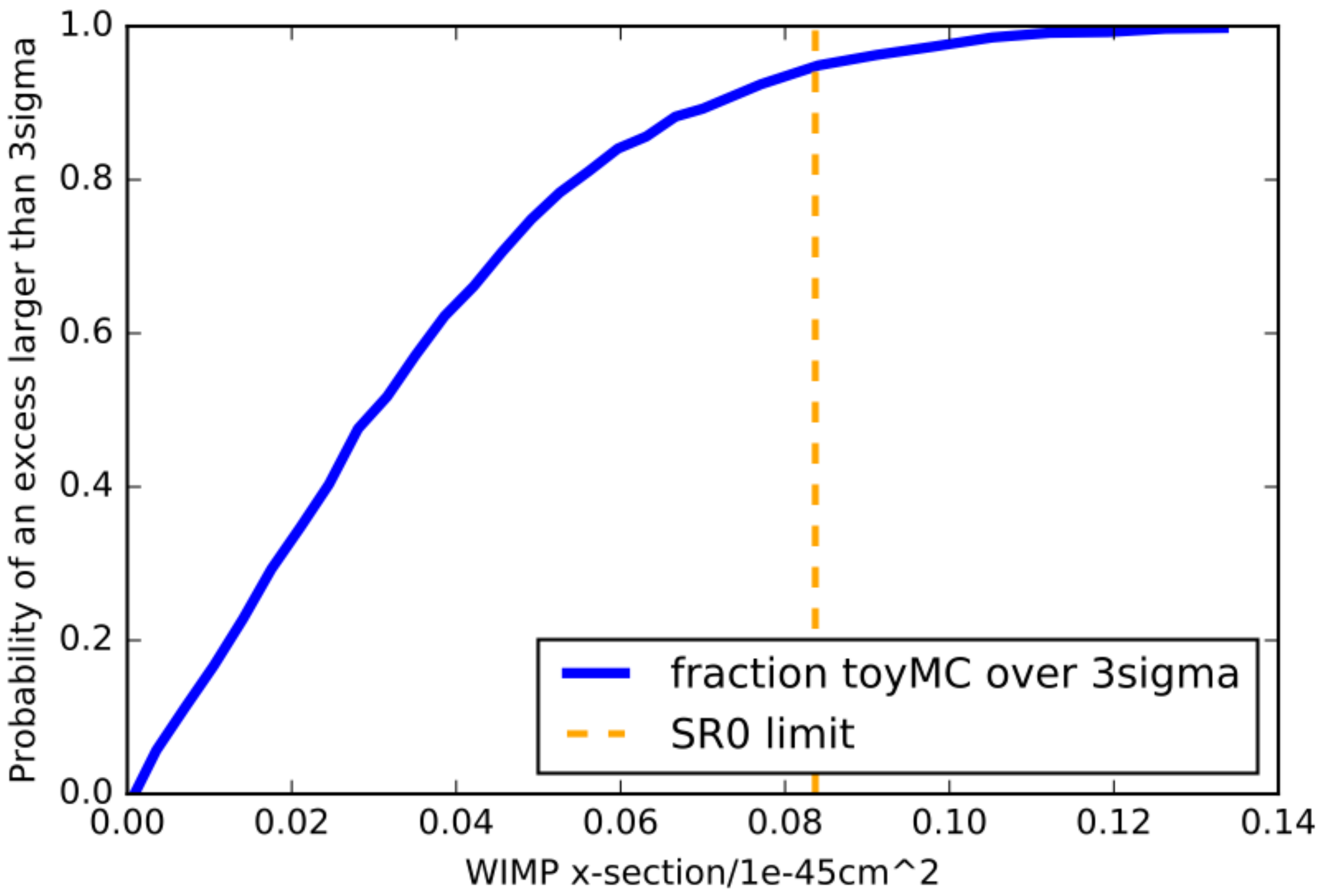


- ATLAS, CMS, LHCb,
- NOvA reports both discovery p-value and CLs value

PCL citation history

**Lux, XENON1T,
PANDAX-II,
ABRACADABRA,
one or two ATLAS,**




Probability of an excess larger than 3sigma
1.0
0.8
0.6
0.4
0.2
0.0
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
WIMP x-section/1e-45cm^2
fraction toyMC over 3sigma
SR0 limit