Expected 95% CL_s Limits for the HNL Displaced-Vertex Search

Asymptotic, exact-Poisson, and toy-MC implementations

August 27, 2026

Abstract

This note documents the three statistical methods implemented in `analysis/scripts/yields/` — `plot_cls_limit_map.py` (asymptotic and exact) and `plot_cls_limit_map_toys.py` (toy MC) — for the expected 95% CL_s exclusion boundary in the $(m_N, |V_{N\mu}|^2)$ plane of the HNL displaced-vertex analysis (CMS EXO-20-009 strategy [4]). All three methods share the same two-bin Poisson counting experiment, the same one-sided profile-likelihood-ratio test statistic, and the same CL_s criterion; they differ only in how the sampling distributions of the test statistic are obtained. For each method the formulae are derived and referenced, and the numerical results are compared in Section 6.

Contents

1 The counting experiment and the CL$_s$ criterion

1.1 Likelihood

For each simulated point (m_N, τ_N) of the signal grid, the analysis predicts signal yields in two statistically independent signal bins of the CMS EXO-20-009 muon-channel selection [4]:

- bin 1 (s_1): the high-mass selection (`exp_yield` column of `yields.csv`),
- bin 2 (s_2): the low-mass ($m_N < 4$ GeV), large-Δ_{2D} selection (`exp_yield_mlt4_d2dgt4`).

At any given mass one of the two yields is typically negligible, but both enter the likelihood in general. Each bin is modelled as a Poisson count with mean $\mu s_i + b$, where μ is the signal-strength modifier ($\mu = 1$ is the nominal signal point) and $b = 1$ expected background event per bin, with no systematic uncertainties:

$$L(\mu) \;=\; \prod_{i=1,2} \mathrm{Pois}(n_i \,|\, \mu s_i + b) \;=\; \prod_{i=1,2} \frac{(\mu s_i + b)^{n_i}}{n_i!}\, e^{-(\mu s_i + b)}. \tag{1}$$

1.2 Test statistic

All three methods use the one-sided profile likelihood ratio of Cowan–Cranmer–Gross–Vitells, Eq. (14) of Ref. [1]:

$$q_\mu \;=\; \begin{cases} -2\ln\lambda(\mu) \;=\; -2\ln\dfrac{L(\mu)}{L(\hat\mu)} & \hat\mu \le \mu \quad \text{(data at or below the tested signal)}, \\ 0 & \hat\mu > \mu \quad \text{(upward fluctuation: no exclusion)}, \end{cases} \tag{2}$$

where $\hat\mu \ge 0$ is the unconditional maximum-likelihood estimator. For the two-bin likelihood (1), $\hat\mu$ follows from $\partial \ln L/\partial\mu = 0$,

$$\sum_i \left[\frac{n_i s_i}{\mu s_i + b} - s_i\right] = 0 \;\Longrightarrow\; A\hat\mu^2 + B\hat\mu + C = 0, \tag{3}$$

$$A = (s_1 + s_2)\, s_1 s_2, \qquad B = b(s_1+s_2)^2 - (n_1+n_2)\, s_1 s_2, \qquad C = b^2(s_1+s_2) - b\,(n_1 s_1 + n_2 s_2), \tag{4}$$

so that

$$\hat\mu = \max\left(0,\ \frac{-B + \sqrt{B^2 - 4AC}}{2A}\right), \qquad \text{or, if } s_1 s_2 = 0, \qquad \hat\mu = \max\left(0,\ \frac{n_1+n_2-2b}{s_1+s_2}\right). \tag{5}$$

(If $C \ge 0$ the likelihood is already falling at $\mu = 0$ and the maximum is at the physical boundary $\hat\mu = 0$.) Substituting Eq. (1) into Eq. (2) gives the closed form used everywhere below:

$$q_\mu(n_1, n_2) \;=\; 2\sum_{i=1,2}\left[(\mu - \hat\mu)\, s_i - n_i \ln\frac{\mu s_i + b}{\hat\mu s_i + b}\right] \qquad (\hat\mu < \mu), \tag{6}$$

with the convention $n_i \ln(\cdot) = 0$ for $n_i = 0$.

1.3 The CL_s upper limit

For an observation $n = (n_1, n_2)$ with $q_\mu^{\mathrm{obs}} = q_\mu(n_1, n_2)$, define the two tail probabilities

$$p_\mu = P\Big(q_\mu \geq q_\mu^{\mathrm{obs}} \,\Big|\, \mu s + b\Big), \qquad 1 - p_b = P\Big(q_\mu \geq q_\mu^{\mathrm{obs}} \,\Big|\, b\Big), \tag{7}$$

under the signal-plus-background and background-only hypotheses, respectively. The CL_s ratio [2, 3] is

$$\mathrm{CL}_s(\mu;\, n) \;=\; \frac{p_\mu}{1 - p_b}, \tag{8}$$

and the 95% CL_s upper limit $\mu_{\mathrm{UL}}(n)$ solves

$$\mathrm{CL}_s(\mu_{\mathrm{UL}};\, n) \;=\; \alpha \;=\; 0.05. \tag{9}$$

$\mathrm{CL}_s(\mu)$ decreases monotonically with μ in the relevant region, so Eq. (9) is solved by bracketing and bisection in all three implementations.

Expected limits and bands. Since no real data are used, the *expected* limit is constructed from the distribution of $\mu_{\mathrm{UL}}(n)$ over the possible background-only outcomes $n \sim \mathrm{Pois}(b) \otimes \mathrm{Pois}(b)$. The median expected limit and the $\pm 1\sigma$, $\pm 2\sigma$ band edges are the 50%, 15.87%/84.13% and 2.275%/97.725% quantiles of that distribution ($+N\sigma$ = upward data fluctuation = weaker limit). The three methods differ only in how the probabilities in Eqs. (7)–(8) and these quantiles are evaluated.

2 Method 1: asymptotic formulae

2.1 Wald approximation and the Asimov data set

The asymptotic implementation uses the formulae of Ref. [1]. In the Wald approximation (their Sec. 3.6, Eqs. (54)–(57)), for a single parameter μ with estimator variance σ^2,

$$q_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}), \qquad F\big(q_\mu \,\big|\, \mu'\big) = \Phi\left(\sqrt{q_\mu} - \frac{\mu - \mu'}{\sigma}\right), \tag{10}$$

i.e. $\sqrt{q_\mu}$ is Gaussian with mean $(\mu - \mu')/\sigma$ and unit width. The variance is defined through the *Asimov data set* — the artificial data for which all estimators equal their true values — here $n_i = b_i$ for the background-only hypothesis. Setting $n_i = b_i$ in Eq. (5) gives $\hat{\mu} = 0$, and the Asimov likelihood ratio (Cowan Sec. 3.2, Eqs. (29)–(31): $q_{\mu,A} = \mu^2/\sigma^2$) is, directly from Eq. (6),

$$\begin{aligned} q_{\mu,A} = -2\ln\frac{L(\mu)}{L(0)}\bigg|_{n_i = b_i} &= -2\sum_i \left[b_i \ln\frac{\mu s_i + b}{b} - \mu s_i\right] \\ &= 2\sum_{i=1,2}\left[\mu s_i - b\,\ln\!\Big(1 + \frac{\mu s_i}{b}\Big)\right], \end{aligned} \tag{11}$$

which is the function $f(\mu)$ in `mu_upper_limit()` of `plot_cls_limit_map.py`. Note the $b \to 0$ limit: $q_{\mu,A} \to 2\mu s_{\mathrm{tot}}$, i.e. the Asimov significance squared is simply twice the expected signal yield scaled by μ.

2.2 Asymptotic CL_s

From Eq. (10), the two tail probabilities of Eq. (7) are (Cowan Eq. (59) and its $\mu' = 0$ analog)

$$p_\mu = 1 - \Phi\Big(\sqrt{q_\mu^{\mathrm{obs}}}\Big), \qquad 1 - p_b = 1 - \Phi\Big(\sqrt{q_\mu^{\mathrm{obs}}} - \frac{\mu}{\sigma}\Big) = \Phi\Big(\sqrt{q_{\mu,A}} - \sqrt{q_\mu^{\mathrm{obs}}}\Big), \tag{12}$$

using $\mu/\sigma = \sqrt{q_{\mu,A}}$. The CL_s ratio [2] is therefore

$$\boxed{\mathrm{CL}_s(\mu) \;=\; \frac{1 - \Phi\Big(\sqrt{q_\mu^{\mathrm{obs}}}\Big)}{\Phi\Big(\sqrt{q_{\mu,A}} - \sqrt{q_\mu^{\mathrm{obs}}}\Big)}} \tag{13}$$

(Ref. [1] presents the CL_{s+b} version, i.e. only the numerator; the use of the same asymptotic distributions in the CL_s ratio is the standard practice referred to in their Sec. 3.8, following Read [2].)

2.3 Median and $\pm N\sigma$ expected thresholds

Under background only, $\sqrt{q_\mu^{\mathrm{obs}}} \sim \mathcal{N}(\sqrt{q_{\mu,A}}, 1)$. The $+N\sigma$ expected band edge (upward data fluctuation, weaker limit) corresponds to a downward fluctuation of the test statistic, $\sqrt{q_\mu^{\mathrm{obs}}} = \sqrt{q_{\mu,A}} - N$. Substituting into $\mathrm{CL}_s = \alpha$ with Eq. (13),

$$1 - \Phi\big(\sqrt{q_{\mu,A}} - N\big) = \alpha\,\Phi(N) \;\;\Longrightarrow\;\; \sqrt{q_{\mu,A}} \;=\; N + \Phi^{-1}(1 - \alpha\,\Phi(N)), \tag{14}$$

which is `q_threshold()` in the code. With $\alpha = 0.05$:

band edge	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
N	-2	-1	0	$+1$	$+2$
$q_{\mu,A}$ threshold	1.106	1.994	3.841	7.438	13.366

(the median value $1.96^2 = 3.84$ is the familiar one; for CL_{s+b} instead of CL_s the same exercise gives $(1.96 - 0.51)^2 \approx 2.69$ — CL_s is the more conservative choice). The expected upper limit for each band edge then solves

$$q_{\mu,A}(\mu_{\mathrm{UL}}) = q_{\mu,A}\text{-threshold}, \tag{15}$$

by bisection on the monotonically increasing function (11).

2.4 Known small-sample limitation

For a nearly background-free observation with $n = 0$ events, the exact Poisson upper limit is 3.0 signal events (Section 3.2), but the asymptotic median limit converges, as $b \to 0$, to

$$\mu s_{\mathrm{tot}} \;=\; \frac{q_{\mu,A}\text{-threshold}}{2} \;=\; \frac{3.84}{2} \;=\; 1.92 \text{ events}, \tag{16}$$

since $q_{\mu,A} \to 2\mu s_{\mathrm{tot}}$. The asymptotic construction therefore *overestimates the sensitivity* (limit too strong) in the low-yield regime — the classic "1.92 vs 3.0 events" failure of the Wald approximation at small counts. This is quantified numerically in Section 6.

3 Method 2: exact Poisson enumeration

3.1 Exact tail probabilities

The exact implementation (`exact_mu_limits()` in `plot_cls_limit_map.py`) evaluates the two tails of Eq. (7) by direct summation over the two-bin outcome space, with the likelihood-ratio ordering of Eq. (6):

$$p_\mu(n;\,\mu) = \sum_{\substack{k_1,k_2\geq 0\\ q_\mu(k_1,k_2)\geq q_\mu(n_1,n_2)}} \mathrm{Pois}(k_1|\mu s_1+b)\,\mathrm{Pois}(k_2|\mu s_2+b), \tag{17}$$

$$1-p_b(n;\,\mu) = \sum_{\substack{k_1,k_2\geq 0\\ q_\mu(k_1,k_2)\geq q_\mu(n_1,n_2)}} \mathrm{Pois}(k_1|b)\,\mathrm{Pois}(k_2|b). \tag{18}$$

The observed outcome itself is included in the sums (the standard convention for discrete distributions; ties at $q_\mu = 0$ are handled with a small relative tolerance). In practice the sums are truncated at $k_i \lesssim \lambda_i + 15\sqrt{\lambda_i}$ ($\lambda_i = \mu s_i + b$), which is exact to machine precision, and $\hat{\mu}$ for every (k_1, k_2) is the exact quadratic solution of Eq. (5). For each outcome n the upper limit $\mu_{\mathrm{UL}}(n)$ then solves Eq. (9) by bisection; CL_s is vectorised over all outcomes simultaneously (`numpy`), so one solve produces the whole expected-limit distribution.

3.2 Closed-form check: $n = (0, 0)$

For the empty observation, $\hat{\mu} = 0$ and Eq. (6) gives $q_\mu(0,0) = 2\mu(s_1+s_2)$. Any other outcome has a strictly smaller q_μ: outcomes with $\hat{\mu}(k) = 0$ satisfy

$$q_\mu(k) = 2\mu(s_1+s_2) - 2\sum_i k_i \ln\Big(1+\frac{\mu s_i}{b}\Big) < q_\mu(0,0), \tag{19}$$

and outcomes with $\hat{\mu}(k) > 0$ are smaller still. Hence only $(0,0)$ contributes to the sums (17)–(18):

$$p_\mu = e^{-(\mu s_1+b)}\,e^{-(\mu s_2+b)}, \qquad 1-p_b = e^{-2b} \implies \mathrm{CL}_s = e^{-\mu(s_1+s_2)}, \tag{20}$$

$$\boxed{\mu_{\mathrm{UL}}(0,0) = \frac{\ln 20}{s_1+s_2} = \frac{2.996}{s_1+s_2}} \tag{21}$$

— the canonical "3.0 events for zero observed events" limit. This is reproduced exactly by the code: for $s_1 = 3$, $s_2 = 0$, $b = 0.01$ it returns $\mu_{\mathrm{UL}} = 0.9986 = 2.996/3$, where the asymptotic formula gives 0.658 (Section 2.4).

3.3 Expected-limit distribution and its discreteness

The expected-limit distribution is the set of values $\mu_{\mathrm{UL}}(n)$ weighted by the background-only probabilities $w(n) = \mathrm{Pois}(n_1|b)\,\mathrm{Pois}(n_2|b)$. Only outcomes with combined weight $> 10^{-9}$ are kept ($\sim 10^2$ outcomes for $b = 1$/bin) and the weights are renormalised. The five band edges are the quantiles 0.02275, 0.15865, 0.5, 0.84135, 0.97725 of this discrete distribution, with linear interpolation between adjacent *distinct* μ_{UL} values on the cumulative-probability axis.

Because the distribution is discrete, the bands are intrinsically asymmetric. For $b = 1$/bin the total count $n_{\mathrm{tot}} = n_1 + n_2$ is a good proxy for the likelihood-ratio ordering (exactly so when $s_1 = s_2$ and $\hat{\mu} = 0$), and its cumulative weights are:

n_{tot}	weight	cumulative	μ_{UL} (exact)†	quantiles landing here
0	0.1353	0.1353	0.4993	-2σ (2.28%)
1	0.2707	0.4060	0.6361	-1σ (15.9%), median (50%)
2	0.2707	0.6767	0.8039	median, $+1\sigma$ (84.1%)
3	0.1804	0.8571	0.9972	$+1\sigma$
4	0.0902	0.9473	1.2068	$+2\sigma$ (97.7%)
5	0.0361	0.9834	1.4236	$+2\sigma$

†for the benchmark $s_1 = s_2 = 3$, $b = 1$/bin; outcomes with equal n_{tot} share the same μ_{UL} here.

Two consequences visible on the limit map:

- the -2σ and -1σ edges nearly coincide: the -2σ quantile falls entirely inside the $n_{\mathrm{tot}} = 0$ outcome (weight 13.5%), and the -1σ quantile only interpolates $\sim 9\%$ of the way toward the next outcome — a $\sim 1\%$ effect on the boundary, invisible on a logarithmic plot;
- the $+1\sigma$ and $+2\sigma$ edges spread over the $n_{\mathrm{tot}} = 2 \dots 5$ outcomes and differ by $\sim$ 15–20% in the boundary lifetime: the band width lives almost entirely on the weak-limit side. This is genuine low-statistics behaviour, which the smooth asymptotic bands cannot reproduce.

4 Method 3: toy Monte Carlo

4.1 Monte Carlo estimators

The toy implementation (`toy_mu_limits()` in `plot_cls_limit_map_toys.py`) replaces the exact sums (17)–(18) by Monte Carlo estimates of the same two tail probabilities:

$$p_\mu(n;\,\mu) \simeq \frac{1}{N_{s+b}} \sum_{t=1}^{N_{s+b}} \Theta\Big(q_\mu(k^{(t)}) - q_\mu(n)\Big)\,, \qquad k^{(t)} \sim \mathrm{Pois}(\mu s_1 + b) \otimes \mathrm{Pois}(\mu s_2 + b), \tag{22}$$

$$1 - p_b(n;\,\mu) \simeq \frac{1}{N_b} \sum_{t=1}^{N_b} \Theta\Big(q_\mu(k^{(t)}) - q_\mu(n)\Big)\,, \qquad k^{(t)} \sim \mathrm{Pois}(b) \otimes \mathrm{Pois}(b), \tag{23}$$

with the same test statistic (6) and exact $\hat{\mu}$ (5). The binomial uncertainty on a tail p with N toys is $\sqrt{p(1-p)/N}$ (a few per cent relative on CL_s near the crossing for $N = 10^4$); it cancels in the ensemble quantiles below.

4.2 Expected limits from pseudo-data

The expected-limit procedure is the toy analogue of Section 3.3:

1. generate *one* background-only pseudo-data ensemble $d^{(r)} = (n_1^{(r)}, n_2^{(r)})$, $r = 1 \dots 10^4$, reused at every μ (a data set is a data set: its q_μ is simply re-evaluated at each μ);
2. at each μ of a scan (41 log-spaced points over $[\mu_{\mathrm{scale}}/25,\, 25\,\mu_{\mathrm{scale}}]$, with μ_{scale} the asymptotic median of Section 2), generate 10^4 fresh $s + b$ toys;
3. for every pseudo-data toy r form $\mathrm{CL}_s^{(r)}(\mu)$ from the toy fractions (22)–(23), where $1 - p_b$ is the rank of toy r in its own ensemble (ties included; outcomes equal within a relative tolerance of 10^{-12} count as the same tie, as in the exact script);
4. find $\mu_{\mathrm{UL}}^{(r)}$ from the $\mathrm{CL}_s^{(r)} = 0.05$ crossing (monotonicity enforced, linear interpolation in $\ln \mu$);
5. the median and $\pm 1\sigma$, $\pm 2\sigma$ expected limits are the *empirical* quantiles of the $\mu_{\mathrm{UL}}^{(r)}$ distribution — the standard toy Brazil-band construction.

All ensembles use deterministic per-grid-point seeds, so results are reproducible independent of parallel scheduling.

Quantile-convention remark. For the discrete outcome distribution of Section 3.3, the empirical (step) quantiles used here differ slightly from the interpolated quantiles used in the exact script: e.g. any quantile between 13.5% and 40.6% returns exactly the $n_{\rm tot} = 1$ value. The per-outcome $\mu_{\rm UL}$ values themselves agree between the two implementations to ~ 0.5–3% (residual μ-grid interpolation in the toys), so the toy band edges are equal to or slightly weaker than the exact ones — visible as a few per cent narrowing of the excluded lifetime range in Table 2.

4.3 Worked example

Figure 1 shows the construction for one example pseudo-data set, $n_{\rm obs} = (1, 1)$, at the benchmark $s_1 = s_2 = 3$, $b = 1$/bin. For $s_1 = s_2$ the test statistic (6) depends only on $n_{\rm tot} = n_1 + n_2$, so both sampling distributions are discrete spikes, one per $n_{\rm tot}$ class (panel a). At $\mu = \mu_{\rm UL}(n_{\rm obs})$ the toy fractions at or above $q_\mu^{\rm obs} = 4.87$ give $p_\mu = 0.032$ under s+b and $1 - p_b = 0.676$ under b-only, hence $\mathrm{CL}_s = 0.047 \simeq 0.05$ (up to Monte Carlo statistics). Panel (b) shows the full $\mathrm{CL}_s(\mu)$ scan for this data set: the 0.05 crossing defines its upper limit, $\mu_{\rm UL} = 0.82$ (exact per-outcome value 0.804; the difference is toy statistics and scan interpolation).

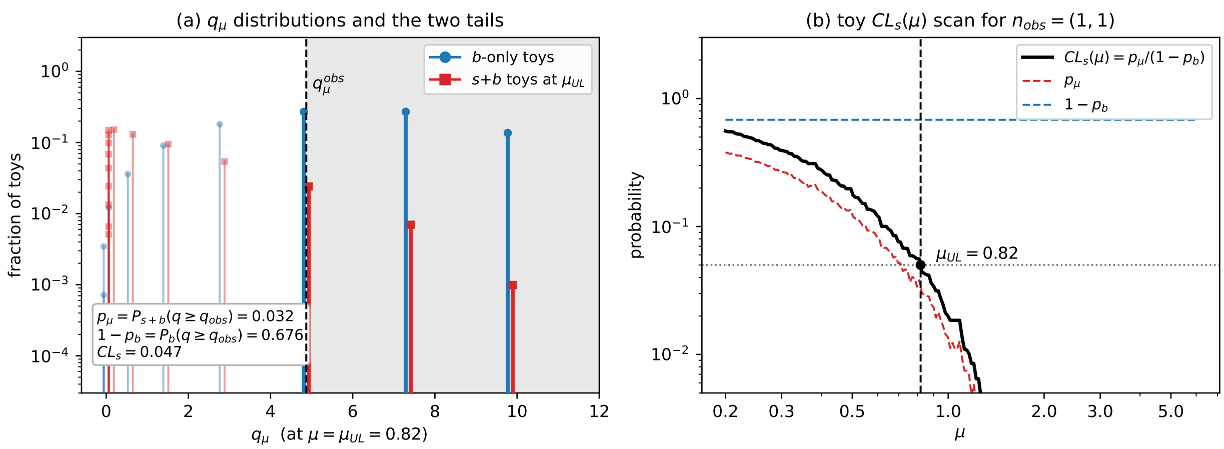

Figure 1: One example pseudo-data set $n_{\rm obs} = (1, 1)$ at $s_1 = s_2 = 3$, $b = 1$/bin. (a) Discrete q_μ distributions under b-only and s+b at $\mu = \mu_{\rm UL}(n_{\rm obs})$; the shaded tail $q_\mu \geq q_\mu^{\rm obs}$ gives the two toy-estimated probabilities entering CL_s. (b) Toy $\mathrm{CL}_s(\mu)$ scan for the same data set, together with its numerator p_μ and denominator $1 - p_b$; the crossing with 0.05 is the per-toy upper limit.

Repeating the scan for all 10^4 pseudo-data sets gives the per-toy upper-limit distribution of Fig. 2: one discrete spike per $n_{\rm tot}$ class, with heights equal to the Poisson($2b$) class probabilities. The toy band edges are the empirical quantiles of this distribution (blue lines) and each lands exactly on a spike. The red dotted lines are the interpolated quantiles of the exact enumeration (Section 3.3), which fall *between* spikes — e.g. the exact median (0.694) interpolates between the $n_{\rm tot} = 1$ and $n_{\rm tot} = 2$ spikes, while the toy median (0.799) sits on the $n_{\rm tot} = 2$ spike. This is the quantile-convention difference of the remark above, made visible. Note also the -2σ edge: toy and exact agree (0.504 vs 0.499) because both fall inside the $n_{\rm tot} = 0$ spike — the discrete band collapse of Section 3.3.

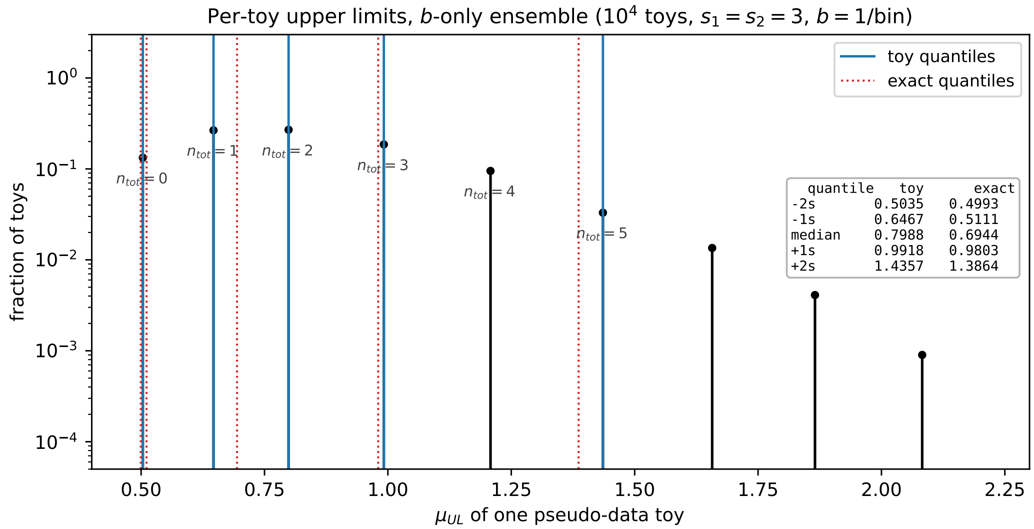

Figure 2: Distribution of the per-toy upper limits over the 10^4-toy background-only pseudo-data ensemble at $s_1 = s_2 = 3$, $b = 1/\text{bin}$. Each spike is one n_{tot} class. Blue solid lines: toy (empirical) quantiles; red dotted lines: exact (interpolated) quantiles, with numerical values in the inset.

5 From $\mu_{\text{UL}}(m_N, \tau_N)$ to the exclusion boundary

The boundary construction is identical for all three methods. It is implemented in `boundary_taus()` and `boundary_polygon()` of `plot_cls_limit_map.py`:

1. at each simulated grid point (m_N, τ_N), compute μ_{UL} for each of the five expected quantiles (s_1, s_2 from `yields.csv`);
2. for each mass, scan $\mu_{\text{UL}}(\tau)$ along the simulated lifetimes and find the crossings of $\mu_{\text{UL}} = 1$ (the model is excluded where $\mu_{\text{UL}} \leq 1$). Crossings are log-log interpolated between adjacent grid lifetimes t_0, t_1 with limits μ_0, μ_1:

$$t_c = t_0 \left(\frac{t_1}{t_0}\right)^{\ln \mu_0 / \ln(\mu_0/\mu_1)}; \tag{24}$$

   the first down-crossing (short τ) and last up-crossing (long τ) bracket the excluded lifetime range; if the model is excluded out to the grid edge, the boundary is clipped at the edge lifetime;
3. convert τ_N to the mixing via $|V_{N\mu}|^2 = \hbar/(\Gamma_N^0\, \tau_N)$, where $\Gamma_N = \Gamma_N^0(m_N)\,|V_{N\mu}|^2$ is the total width at unit mixing (`hnl_v2`);
4. connect the boundary points of all masses with straight lines into one closed polygon per quantile, drawn CMS Fig. 9 style: nested fills (95% yellow, 68% green, white hollow core) and a dashed median outline.

6 Numerical comparison

6.1 Benchmark points

Table 1 compares the three methods at two fixed yield points: a low-statistics case ($s_1 = 3$, $s_2 = 0$, $b = 0.01$, nearly the zero-background $n = 0$ situation of Section 3.2) and a moderate-statistics two-bin case ($s_1 = s_2 = 3,\ b = 1/\text{bin}$).

Table 1: Expected μ_{UL} (five quantiles) from the three methods.

$s_1 = 3,\ s_2 = 0,\ b = 0.01$		(exact answer: $\ln 20/3 = 0.9986$, degenerate band)			
method	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
asymptotic	0.198	0.348	0.658	1.259	2.249
exact			0.9986 (all five quantiles identical)		
toy MC			1.0051 (all five quantiles identical)		
$s_1 = s_2 = 3,\ b = 1/\text{bin}$					
method	-2σ	-1σ	median	$+1\sigma$	$+2\sigma$
asymptotic	0.313	0.452	0.696	1.108	1.720
exact	0.499	0.511	0.694	0.980	1.386
toy MC	0.504	0.647	0.799	0.992	1.436

The low-statistics case shows the asymptotic failure of Section 2.4 directly: the asymptotic median (0.658) is 34% too strong, and its band is symmetric and smooth where the true band is degenerate (with $b = 0.01$, 98% of all outcomes are $(0, 0)$). Exact and toy agree with the analytic answer to 0.7%. In the two-bin case the three medians agree at the few-per-cent level (0.696 / 0.694 / 0.799); the exact and toy band shapes show the discrete asymmetry of Section 3.3, and the toy median sits one outcome higher than the exact one because of the quantile convention (empirical step vs interpolated quantiles, Section 4).

6.2 Exclusion reach

Table 2 compares the median expected excluded lifetime ranges on the full signal grid, and Fig. 3 shows the two maps.

Table 2: Median expected excluded τ_N range [ps] per mass ($\mu_{\text{UL}} = 1$ boundary, log-log interpolated). At $m_N = 12$ GeV the model is excluded only by the $-1\sigma/-2\sigma$ band edges in the exact (3.1–10 ps) and toy (3.0–10.5 ps) calculations.

m_N [GeV]	asymptotic	exact	toy MC
1	0.2 – 9490	0.2 – 9350	0.2 – 8710
2	0.48 – 2630	0.484 – 2590	0.492 – 2440
3	0.89 – 1070	0.894 – 1040	0.912 – 995
5	0.44 – 313	0.442 – 311	0.448 – 294
7	0.56 – 128	0.562 – 128	0.576 – 121
10	1.45 – 33.6	1.49 – 32.4	1.53 – 31.7
12	4.16 – 7	—	—

The pattern is exactly as expected from the benchmarks:

- **asymptotic vs exact:** the exact limits are 2–8% weaker in μ_{UL} everywhere, so the excluded region shrinks slightly. The most visible change is $m_N = 12$ GeV dropping out of the median exclusion — the asymptotic formula had it barely excluded (4.2–7 ps) while the exact

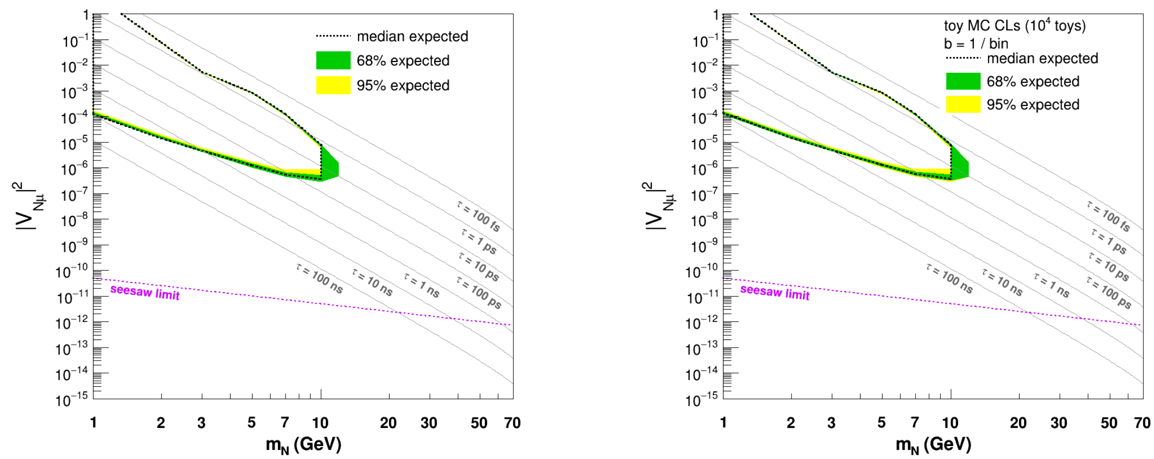

Figure 3: Expected 95% CL$_s$ exclusion boundaries in $(m_N, |V_{N\mu}|^2)$ from the exact Poisson enumeration (left) and the toy-MC method (right). Dashed: median expected; green/yellow: 68%/95% expected bands; grey dashed: constant-τ_N references; violet: seesaw line $|V|^2 = m_\nu/m_N$ for $m_\nu = 0.05$ eV.

calculation places it at the sensitivity edge, excluded only by the downward-fluctuation band edges;

- **exact vs toy:** agreement to a few per cent in the boundary lifetimes; the toy boundary is marginally weaker, entirely accounted for by the step-quantile convention and toy statistics;
- all three methods agree that the sensitivity vanishes for $m_N \geq 13$ GeV and that the strongest reach is at $m_N \simeq 1$ GeV, probing $|V_{N\mu}|^2 \sim 10^{-6}$–$10^{-4}$ at $\tau_N \sim 0.2$ ps–10 ns.

6.3 Implementation map

method	function	file
asymptotic	`mu_upper_limit()`, `q_threshold()`	`plot_cls_limit_map.py`
exact	`exact_mu_limits()`, `compute_limits()`	`plot_cls_limit_map.py`
toy MC	`toy_mu_limits()`, `compute_limits()`	`plot_cls_limit_map_toys.py`
toy examples	`cls_scan_one_toy()`, figures	`plot_toy_examples.py`
boundary/plot	`boundary_taus()`, `boundary_polygon()`	`plot_cls_limit_map.py`

Exact results are cached in `exact_cls_cache.csv`, toy results in `toy_cls_cache.csv` (keyed by yields, b, and toy settings), so the plots can be regenerated without recomputing the limit grid.

References

[1] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554 [Erratum: Eur. Phys. J. C **73** (2013) 2501], arXiv:1007.1727 [physics.data-an]. Test statistic: Sec. 2.4, Eq. (14); Asimov data: Sec. 3.2, Eqs. (29)–(31); Wald approximation, p-values, upper limits: Sec. 3.6, Eqs. (54)–(61); median and $\pm N\sigma$ bands: Sec. 4.3, Eqs. (88)–(89); CL$_s$ remark: Sec. 3.8.

[2] A. L. Read, *Presentation of search results: the CL$_s$ technique*, J. Phys. G **28** (2002) 2693.

[3] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. A **434** (1999) 435, arXiv:hep-ex/9902006.

[4] CMS Collaboration, *Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **07** (2022) 122, arXiv:2201.05578 [hep-ex].