Appendix: Numerical Examples

Tool Reference

RST Reference Number: RST24MD19.01

Date of Publication: 09/22/2025

Recommended Citation: U.S. Food and Drug Administration. (2025). *DxGoals: A Software Tool for Determining, Visualizing, and Analyzing Performance Goals for Diagnostic Test Classification Accuracy* (RST24MD19.01). https://cdrh-rst.fda.gov/dxgoals-software-tool-determining-visualizing-and-analyzing-performance-goals-diagnostic-test

For more information

Catalog of Regulatory Science Tools to Help Assess New Medical Devices

Disclaimer

About the Catalog of Regulatory Science Tools

The enclosed tool is part of the Catalog of Regulatory Science Tools, which provides a peer-reviewed resource for stakeholders to use where standards and qualified Medical Device Development Tools (MDDTs) do not yet exist. These tools do not replace FDA-recognized standards or MDDTs. This catalog collates a variety of regulatory science tools that the FDA's Center for Devices and Radiological Health's (CDRH) Office of Science and Engineering Labs (OSEL) developed. These tools use the most innovative science to support medical device development and patient access to safe and effective medical devices. If you are considering using a tool from this catalog in your marketing submissions, note that these tools have not been qualified as Medical Device Development Tools and the FDA has not evaluated the suitability of these tools within any specific context of use. You may request feedback or meetings for medical device submissions as part of the Q-Submission Program.

For more information about the Catalog of Regulatory Science Tools, email RST_CDRH@fda.hhs.gov.

Appendix: Numerical Examples

1. Penicillin Allergy Skin Test to **Rule Out** Risk of Allergic Reactions in Patients Labeled as Penicillin Allergic. Among patients who have historically been labeled as penicillin allergic, suppose the prevalence (pre-test probability) of hypersensitivity reaction (HSR) to amoxicillin is p= 0.051 (5.1%) and that a clinically acceptable risk threshold for ruling out HSR is $0.005\ (0.5\%) = cNPV^* = 1 - NPV^*$, where NPV is the negative predictive value. With the inputs $p = 0.051$ and $cNPV^* = 0.005$,, DxGoals computes that a test negative result for ruling out HSR should satisfy $NLR < NLR^* = 0.094$, where $NLR = FNF/TNF = (1 - Se)/Sp$ is the negative likelihood ratio, $Se =$ sensitivity, and $Sp =$ specificity. Pairs of (Se, Sp) meeting this goal satisfy $Se > 0.906 + 0.094(1 - Sp)$, which when both sides are equated is known as the *negative likelihood ratio line* (Biggerstaff 2000) and include the pairs $(Se, Sp) > (0.95, 0.54)$.
2. **Rule-In** of Surgical Work-Up for Early Detection of Ovarian Cancer. A biomarker test positive result for ovarian cancer may lead to transvaginal ultrasound, upon which if a cyst is found, may lead to referral to pelvic mass surgery, which may be unnecessary if the cyst is benign and lead to co-morbidity. A risk threshold of $PPV^* = 0.1\ (10\%)$ for ruling in surgical work-up (i.e., 10 surgeries for every 1 ovarian cancer found) has been proposed (Skates et al. 2013). Among 40 to 64-year-old women, the annual incidence rate of ovarian cancer is 0.158% (SEER 2023), which may be revised up to $p = 0.00159\ (0.159\%)$ to account for women without ovarian cancer who have undergone bilateral oophorectomy (Erickson et al. 2022). With the inputs $p = 0.00159$ and $PPV^* = 0.1$, DxGoals computes that a test positive result for ruling in surgical work-up for ovarian cancer should satisfy $PLR > PLR^* = 69.77$, where $PLR = TPF/FPF = Se/(1 - Sp)$ is the positive likelihood ratio. Pairs of (Se, Sp) meeting this goal satisfy $Se > 69.77 \times (1 - Sp)$, which when both sides are equated is known as the *positive likelihood ratio line* (Biggerstaff 2000) and include the pairs $(Se, Sp) > (0.6977, 0.99)$.
3. Human Papilloma Virus (HPV) Test to **Rule in** and **Rule Out** 5-Year Risk of Cervical Cancer. HPV tests are used to screen asymptomatic women for cervical squamous intraepithelial neoplasia stage 3 or cervical cancer (CIN3+). Suppose the 5-year cumulative incidence rate of CIN3+ in asymptomatic women is $0.005\ (0.5\%)$. Based loosely on guidelines (Katki et al 2013; Perkins et al. 2020), suppose women are recommended to return in 5-years when their 5-year risk of CIN3+ is 0.25% or less and referred to immediate colposcopy when the risk is 1% or greater. With the inputs $cNPV^* = 0.0025, p = 0.005, PPV^* = 0.01$, DxGoals computes that an HPV test positive result for ruling-in immediate colposcopy should satisfy $PLR > PLR^* = 2.01$ and an HPV test negative result for ruling out colposcopy in favor of return in 5 years should satisfy $NLR < NLR^* = 0.499$. The NLR and PLR lines are therefore $Se = 0.501 + 0.499 \times (1 - Sp)$ and $Se = 5.91 \times (1 - Sp)$, which DxGoals plots on the *likelihood ratio (LR) graph* (**Figure 2**) (Biggerstaff 2000). For the counts of true negative, false negative, false positive, and true positive test results $(x_{00}, x_{01}, x_{10}, x_{11}) = (975, 3, 597, 32)$, DxGoals displays analysis results (**Figure 3**) and visualizes them on the LR graph (**Figure 2**). The HPV test meets both of the NLR and PLR goals with statistical significance. The intersection of the PLR and NLR lines determine the sometimes-preferred but more stringent FPF and TPF goals $FPF^* = 0.332$ and $TPF^* = 0.667$, which are, respectively, not met and met for the given dataset. In More, DxGoals produces an extra plot of all risk stratification triples $(cNPV^*, p, PPV^*)$ that confer the same NLR and PLR goals as those input, in this case $(cNPV^*, p, PPV^*) = (0.0025, 0.005, 0.01)$.

(**Figure 4**).

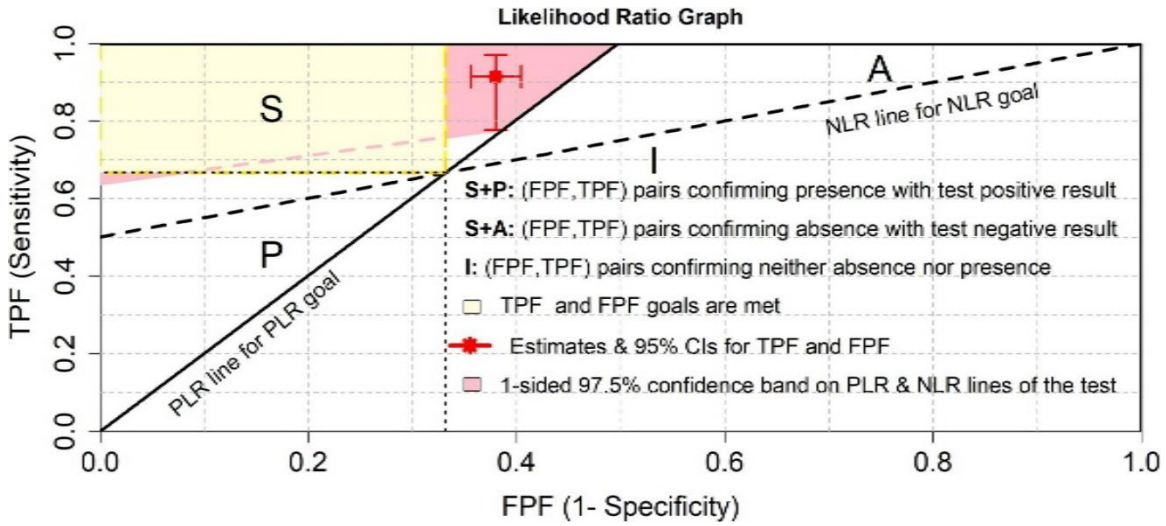

Figure 2. Likelihood ratio graph of the NLR and PLR lines corresponding to the NLR and PLR goals implied by risk stratification $(cNPV^*, p, PPV^*) = (0.0025, 0.005, 0.01)$ and visualization of whether these goals were met with the data $(x_{00}, x_{01}, x_{10}, x_{11}) = (975,3,597,32)$. Region S consists of the pairs of (FPF, TPF) that meet the NLR and PLR goals. The [pink] shaded quadrilateral is the 1-sided, non-simultaneous 97.5% pointwise confidence band for the PLR and NLR lines of the test. The quadrilateral is subsumed within region S, indicating the PLR and NLR goals are both met with statistical significance. The [yellow] rectangular region is a subset of S that consists of pairs of (FPF, TPF) that meet the sometimes-preferred independent goals for FPF and TPF (i.e., goals for sensitivity $= TPF$ and specificity $= 1 - FPF$), indicated by the right and bottom borders of the rectangle. The goal for FPF was not met because the [red] horizontal error bar depicting the 95% confidence interval (CI) for FPF $(0.356,0.404)$ lies to the right of the right border. The goal of TPF was met because the [red] vertical error bar depicting the 95% CI for TPF $(0.776,0.97)$ lies above the bottom border.

The NLR performance goal is NLR ≤ 0.499
Your estimate of NLR is $\widehat{NLR}$ = 0.138
95% confidence interval on NLR is (0.048 , 0.361)
Did you data meet the NLR goal with statistical significance? Yes

The PLR performance goal is PLR ≥ 2.01
Your estimate of PLR is $\widehat{PLR}$ = 2.407
95% confidence interval of PLR is (2.028 , 2.645)
Did you data meet the PLR goal with statistical significance? Yes

Specificity (Sp) goal is Sp ≥ 0.668
Your estimate of Sp is $\widehat{Sp}$ = 0.62
95% confidence interval on Sp is (0.596 , 0.644)
Did your data meet the Sp goal with statistical significance? No

Sensitivity (Se) goal is Se ≥ 0.667
Your estimate of Se is $\widehat{Se}$ = 0.914
95% confidence interval on Se is (0.776 , 0.97)
Did your data meet the Se goal with statistical significance? Yes

Figure 3. DxGoals results for HPV test data $(x_{00}, x_{01}, x_{10}, x_{11}) = (975,3,597,32)$ given risk stratification input $(cNPV^*, p, PPV^*)) = (0.0025, 0.005, 0.01)$.

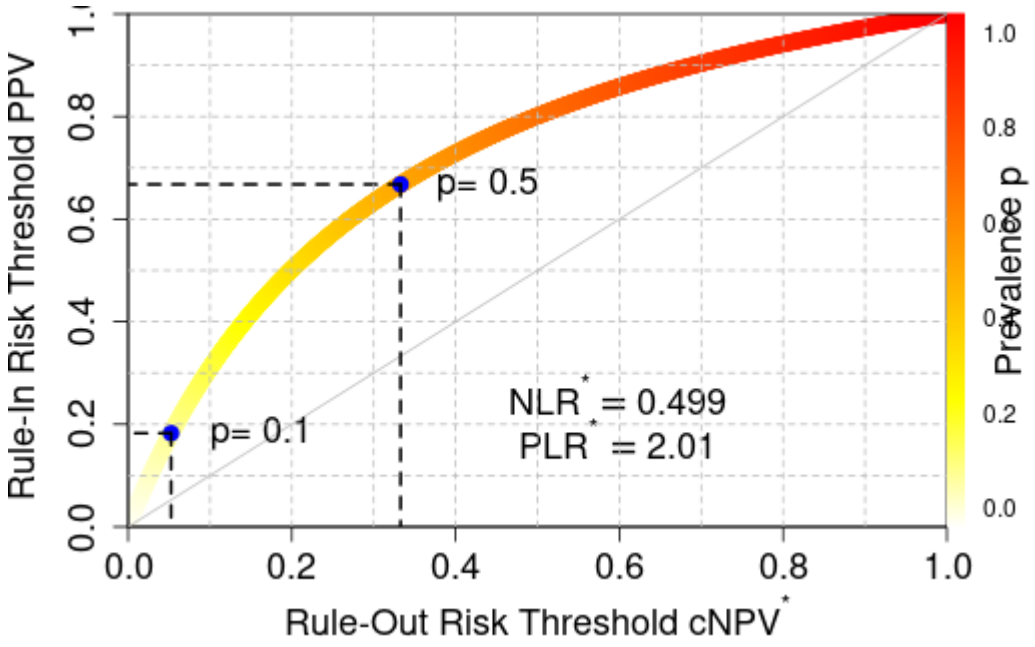

Figure 4. All risk stratification triples that confer NLR and PLR goals $NLR^* = 0.499$ and $PLR^* = 2.01$, including $(cNPV^*, p, PPV^*) = (0.0025, 0.005, 0.01)$.

4. Hypothesis Test of Non-inferiority in Risk Difference for 2 HPV Tests.

Test B is defined as non-inferior to Test A in rule-out risk difference $p - cNPV$ if $p - cNPV^B > \gamma_0 \times (p - cNPV^A)$, where γ_0 is a non-inferiority (NI) margin. This inequality holds approximately when $1 - NLR^B > \gamma_0 \times (1 - NLR^A)$, i.e., $NLR^B < 1 - \gamma_0 + \gamma_0 \times NLR^A$ (Pennello 2021).

Test B is defined as non-inferior to Test A in rule-in risk difference $PPV - p$ if $PPV^B - p > \gamma_1 \times (PPV^A - p)$, where γ_1 is an NI margin. This inequality holds approximately when $PLR^B - 1 > \gamma_1 \times (PLR^A - 1)$ i.e., $PLR^B > 1 - \gamma_1 + \gamma_1 \times PLR^A$ (Pennello 2021).

Suppose $\gamma_0 = \gamma_1 = 0.95$. Consider hypothetical data on two HPV tests A and B for CIN3+ (Table). For tests A and B, DxGoals calculates PLR estimates $\widehat{PLR}^A = 8$ and $\widehat{PLR}^B = 5.5$ and NLR estimates $\widehat{NLR}^A = 0.222$ and $\widehat{NLR}^B = 0.206$. $\hat{R}_0 = \left(1 - \widehat{NLR}^A\right)^{-1}\left(1 - \widehat{NLR}^B\right) = (1 - 0.222)^{-1}(1 - 0.206) = 1.021$ has 95% CI $(0.951, 1.097)$, whose lower limit $0.951 > 0.95$ indicates that test B is non-inferior to test A in ruling out CIN3+. $\hat{R}_1 = \left(\widehat{PLR}^A - 1\right)^{-1}\left(\widehat{PLR}^B - 1\right) = (5.5 - 1)^{-1}(8 - 1) = 0.643$ has 95% CI $(0.578, 0.715)$, whose upper limit $0.715 < 1$ indicates that test A is superior to test B in ruling in CIN3+. DxGoals displays these results (Figure 5).

Table. Comparative data for HPV tests A and B.

Test B	Test A		Total
	A = 0	A = 1	
Nondiseased subjects (D = 0)			
B = 0	6912	432	7344
B = 1	864	432	1296
Total	7776	864	8640
Diseased subjects (D = 1)			
B = 0	27	36	63
B = 1	45	252	297
Total	72	288	360

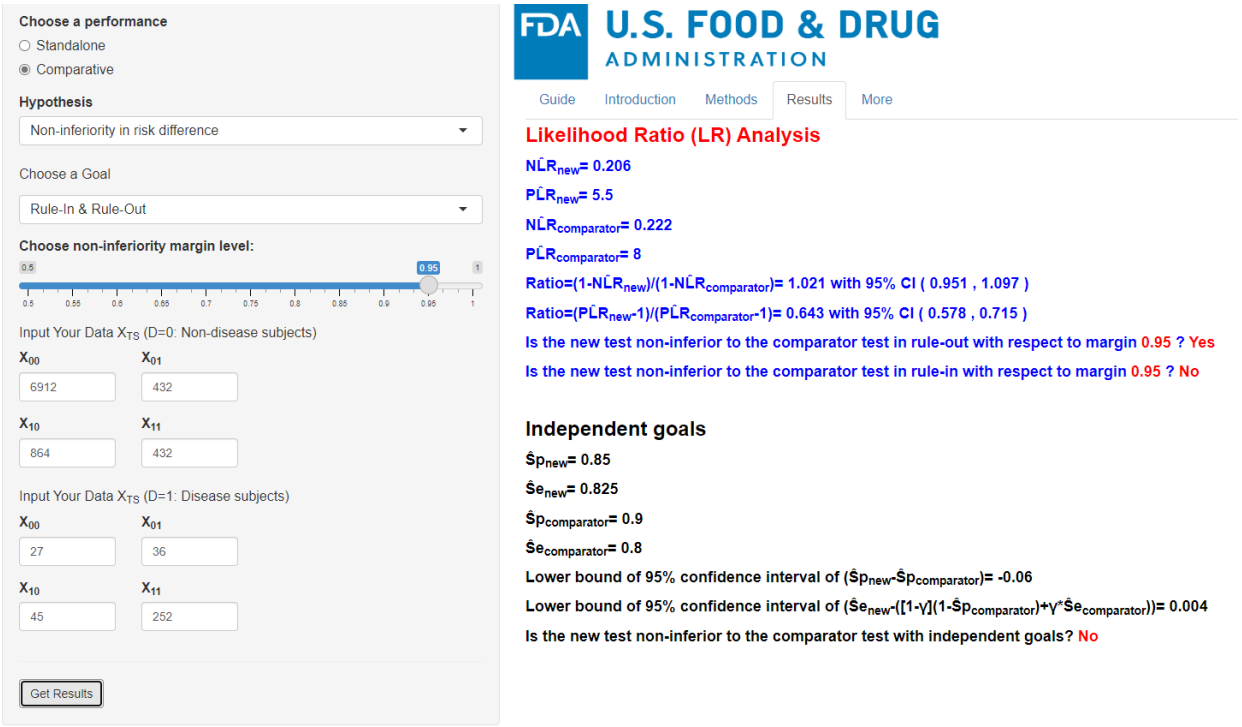

Figure 5. DxGoals analysis of data in Table for hypothesis test of non-inferiority in risk difference with respect to non-inferiority margins of 0.95 for both rule-in and rule-out of CIN3+.