



In Yucub village, located in Ethiopia's Somali region, beneficiaries gathered for a cash distribution organized by WFP.

© WFP/Michael Tewelde



World Food Programme

SAVING
LIVES
CHANGING
LIVES

ANALYSIS, PLANNING & PERFORMANCE

TARGETING ADVISORY—Issue 3, 2025

Using ROC Curves for Targeting Analysis

SUMMARY

This advisory note guides how Receiver Operating Characteristics (ROC) curves can be used to improve targeting design in humanitarian programmes.

The Receiver Operating Characteristics (ROC) curve is a visual tool that shows us the targeting model's performance across all possible eligibility thresholds.

ROC analysis offers a systematic way to evaluate the critical trade-offs between errors of inclusion (giving assistance to those who don't need it) and exclusion (failing to provide aid to those who do need it). This is a crucial consideration for any humanitarian programme that must make difficult decisions with limited resources.

This is not just a technical exercise; it's a way to visualize the ethical and practical choices faced in targeting design.

INTRODUCTION

Humanitarian programmes face the challenge of reaching the most vulnerable households with limited resources. Targeting design inevitably produces two types of errors: **exclusion errors**, people who are in need of assistance but are not included in the targeted group, and **inclusion errors**, programme beneficiaries who do not need assistance but are identified as in need based on the targeting method chosen. The relative importance of these errors depends on the programme context.¹

Targeting analysts are often faced with the challenge of selecting an eligibility threshold that strikes the right balance between minimizing errors of inclusion and exclusion.

The ROC analysis can be utilized for targeting analysis, providing a transparent statistical tool for visualizing and quantifying these design-related trade-offs. **This method is applicable whenever a data-driven model is used for targeting.**

ROC curves help to understand how to choose eligibility thresholds by showing the trade-off between inclusion and exclusion errors at different cut-points. Each point on the curve corresponds to a different probability threshold of being food insecure, displayed according to the targeting design's ability to correctly differentiate food secure vs. food insecure households.

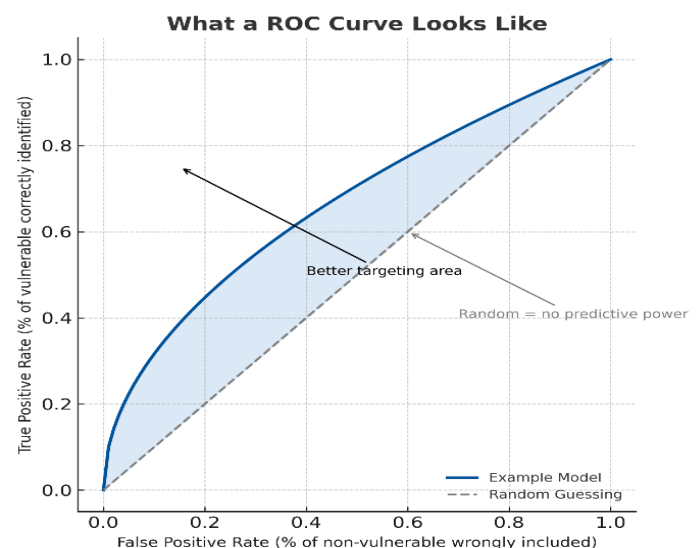
This note emphasizes that selecting a threshold (cut-off point) is not a mechanical task; it is a fundamental programmatic decision. This means there is no single, perfect solution provided by the data; instead, the choice of where to set the eligibility bar must reflect the programme's specific objectives and humanitarian principles, as well as budget constraints.

METHODOLOGY

The design phase of a data-driven targeting process follows a clear, structured process that is consistent across operations. A key step is the establishment of a vulnerability framework, where a clear and measurable definition of “vulnerable” is established, such as a household being food-insecure. This creates a simple yes/no category that the model will predict. Relevant households' characteristics are then identified and used to generate predicted probabilities through statistical models. Next, predicted probabilities give each household a score indicating its likelihood of being vulnerable.

The ROC analysis can then be performed to evaluate the precision of the model across different inclusion-exclusion error thresholds. These curves are like maps that show the range of possible outcomes at different eligibility thresholds, illustrating the model's overall predictive power. The optimal cut-point can be identified using criteria such as Youden's J statistics or program-specific trade-offs between errors.

Figure 1 - Example of a ROC Curve



¹ For more information about targeting errors, you can visit the VAM Resource Center:

<https://docs.wfp.org/api/documents/WFP-0000122035/download/> // <https://docs.wfp.org/api/documents/WFP-0000167583/download/>

EXPLAINING THE ROC CURVE

Think of a targeting model as a metal detector at an airport. Every household passes through, and instead of a simple “yes/no” answer, the detector gives a beep of a certain strength (a score between 0 and 1). A weak beep (0.1) means the household is probably not vulnerable, while a strong beep (0.9) means they likely are.

The **ROC curve** is a picture that shows all the possible ways you can set the detector’s sensitivity. If you set the bar very low, almost everyone will trigger a beep—you’ll catch all the vulnerable households (high recall) but also incorrectly include many non-vulnerable ones (high false positives). If you set the bar very high, only the strongest beeps will trigger—you’ll miss some vulnerable households (exclusion errors) and have fewer non-vulnerable ones (low inclusion errors). The curve visualizes this entire range of trade-offs.

A better model pushes the curve higher, meaning it is more effective at separating the vulnerable from the non-vulnerable. We don't use the ROC curve to make a single decision, but rather to see the entire landscape of trade-offs. This allows us to choose a sensitivity (threshold) that makes the most sense for our programme's goals.

UNDERSTANDING PROBABILITY IN TARGETING MODELS

When building a data-driven targeting model, the output is not a simple “yes” or “no” about whether a household is vulnerable. Instead, the model produces a probability score between 0 and 1. This score represents the likelihood that the household is vulnerable, given the observed characteristics. For example, a score of 0.8 means the household has an 80% chance of being vulnerable, while a score of 0.2 means there is only a 20% chance. These probabilities allow the ranking of households from most to least likely to be vulnerable and then apply a threshold to decide who should be eligible for assistance.

The analysis uses proxy indicators such as household size, dependency ratio, education level, or housing conditions—factors that are observable and relatively stable over time – and not outcome indicators as per corporate guidance.² The aim is to use structural and demographic proxies that help anticipate who is at risk, rather than reproduce the outcomes we are trying to change.

These proxy variables can then be combined in different ways. One common approach is a

scorecard, where each variable is assigned, a weight based on how strongly it predicts vulnerability. Adding up these weights gives each household a final score, which corresponds to a probability of being vulnerable. But scorecards are not the only option. Statistical models such as logistic regression, or machine-learning methods like Random Forests, also generate probability scores.

ROC curves can be applied to any targeting approach that produces a probability or continuous score, such as logistic regression, random forests, or scorecards. In these cases, predicted probabilities (or scores) can be compared against observed outcomes to evaluate performance. However, ROC analysis is not applicable to purely categorical targeting methods unless they are first converted into a scoring system. Once a scoring or probability framework is established, a threshold can then be chosen—guided by ROC curves and other analyses—to balance inclusion and exclusion errors in line with operational and ethical priorities.

² <https://docs.wfp.org/api/documents/WFP-0000169093/download/>

FROM PROBABILITIES TO ROC CURVES

The figure below (Figure 2) shows how each household is assigned a probability of being vulnerable, and how different thresholds change the probability of a household being selected. A low threshold (orange line) includes more households in the programme (with the chance of including households that are not necessarily vulnerable), while a high threshold (green line) is stricter, reducing the chance of better-off households being included but also increasing the possibility of vulnerable people being excluded.

The ROC curve takes this simple threshold logic and stretches it across *all possible thresholds*. Instead of just looking at one cut-off point, it plots the full spectrum. In turn it shows, for each possible threshold, how many vulnerable households are correctly included (recall) versus how many non-vulnerable households are wrongly included (false positives/inclusion error). In other words, the ROC curve is simply a map of what happens when you slide the threshold up and down.

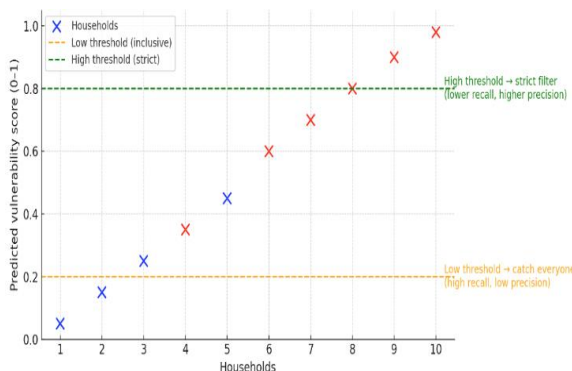


Figure 2 - Thresholds in Probability-based Targeting
Red X – Vulnerable, Blue X – Not Vulnerable

The application of ROC analysis to targeting studies reveals insights into model performance and the

consequences of our choices. For instance, a statistically balanced threshold might offer a good compromise between inclusion and exclusion errors. However, a programmatic choice to align with a pre-determined caseload can significantly shift this balance, prioritizing a wider reach over perfect accuracy.

HOW TO READ A ROC GRAPH

Figure 3 below, shows an example where two models (orange and blue curves) are compared:

- **The Baseline:** The diagonal dashed line represents a model that is no better than random guessing. Any good model's curve will be above this line.
- **The Axes:**
 - The vertical axis (y-axis) shows the share of all vulnerable people correctly identified, also known as "Recall or Sensitivity." A higher value is better.
 - The horizontal axis (x-axis) shows the share of non-vulnerable people incorrectly identified as vulnerable, also known as a "False Positive Rate" or the "specificity". A lower value is better.
- **The Curves:** Each colored curve represents a different targeting model. The closer a curve is to the top-left corner of the graph, the better the model performs.

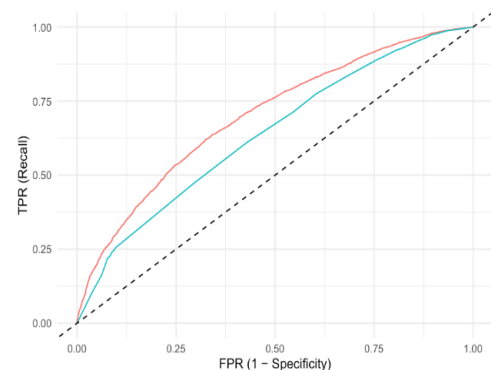


Figure 3 - ROC Comparison

PREDICTING VULNERABILITY TO MALNUTRITION IN MALAWI – A REAL-LIFE EXAMPLE

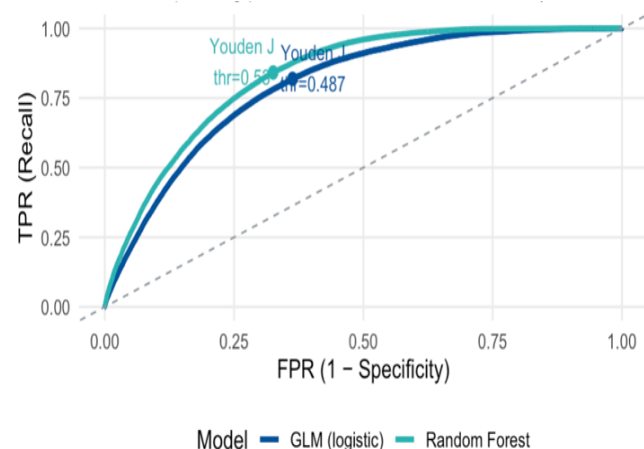
The Malawi Country Office (CO) requested a targeting analysis to understand whether Nutrition Rehabilitation Units (NRUs) data could reliably identify households with nutritional stress.

The assumption was simple: households who reached an NRU were already facing nutritional stress. Identifying the household characteristics most predictive of nutritional stress would enable the CO to more effectively target households in need of nutritional assistance.

Two models were compared: a logistic regression, which provided a transparent and easy-to-explain statistical approach, and a Random Forest, which used many decision trees combined to capture more complex patterns. Both were trained on the same data and then evaluated to see how well they could distinguish between households that did and did not have an NRU admission.

The results showed that both models performed well, with the Random Forest achieving slightly higher accuracy.

Figure 4 - Model Comparison



FURTHER READINGS ON PROBABILISTIC METHODS AND ROC CURVES

- Aiken et al. (2022) – “Machine Learning and Mobile Phone Data Can Improve the Targeting of Humanitarian Assistance” (NBER Working Paper)
- Noriega-Campero et al. (2020) – “Algorithmic targeting of social policies: Fairness, accuracy, and distributed governance”
- Kalaycıoğlu et al. (2023) – “Using machine learning to identify predictors of social vulnerability in the event of a hazard: Istanbul case study”
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves.

CALCULATIONS AND GUIDELINES USING R

Guide to Credit Scoring in R: A practical guide (dated ~2009) that covers calculating ROC curves for logistic regression and random forest models, using the ROCR R package (with code examples).

Caret Package Documentation: Official documentation for the caret R package includes instructions on computing ROC curves and AUC.



ASSESSMENT AND TARGETING UNIT (APP-FA)

ANALYSIS, PLANNING AND PERFORMANCE DIVISION

Contact us:

wfp_assessment_targeting@wfp.org

World Food Programme

Via Cesare Giulio Viola 68/70
00148 Rome, Italy - T +39 06 65131
wfp.org

Targeting Team:

Alberto Gualtieri, Cagri Cebisli,
Edgar Wabyona, Eve Chaliflour,
Federica Esu, Hajar Anbar