Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Go Decision Guided Cutoff Determination for Companion Diagnostics Development

Hong Wang
Sanofi
Shuguang Huang
Stat4ward LLC
Yiding Zhang
Sanofi
Xiaohui Chen
Stat4ward LLC
Shen Zhang
Stat4ward LLC
Wenting Wang
Wenting.Wang2@sanofi.com
Sanofi

Method Article

Keywords: CDx, cutoff, 3ODM, ROC, GDGM, GPGM

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11447564/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Go Decision Guided Cutoff Determination for Companion Diagnostics Development

Hong Wang [a]*, Shuguang Huang[b], Yiding Zhang[c], Xiaohui Chen[b], Shen Zhang[a] and Wenting Wang[a]

[a]*Sanofi, Translational medicine and research statistics, Evidence Generation and Decision Science, 350 Water St, Cambridge, MA 02141;*
[b]*Stat4ward, 2 Edgemoor Lane, Pittsburgh, PA 15238;*

[c]*Sanofi, Oncology Early Development, Evidence Generation and Decision Science, 350 Water St, Cambridge, MA 02141*

CONTACT Wenting Wang Wenting.Wang2@sanofi.com; Sanofi, Translational medicine and research statistics, Evidence Generation and Decision Science, 350 Water St, Cambridge, MA 02141

Go Decision Guided Cutoff Determination for Companion Diagnostics Development

A patient-selection CDx classifies patients into binary results. When the assay is quantitative or semi-quantitative, setting the cutoff faces the challenge of balancing multiple key factors including analytical performance, clinical efficacy, and prevalence. In addition, it also faces the challenge of limited sample size since the cutoff decision typically has to be made based on phase 1 or phase 2 studies. The traditional ROC method determines the cutoff solely based on the mathematical criteria without incorporating clinical considerations. Based upon the established Go/No-Go clinical decision-making framework, the Go decision guided method (GDGM) is proposed to determine the optimal cutoff that maximizes the target population (i.e. high prevalence for biomarker positive patients) while ensuring clinical efficacy. As an extension to the GDGM method, the Go probability guided method (GPGM) uses a Bayesian approach to gate Go probability with a pre-specified sample size to overcome the issue of GDGM that the sample size of the biomarker-positive group decreases as the cutoff increases. Simulations are conducted to evaluate and compare the performance of the three methods.

Keywords: CDx, cutoff, 3ODM, ROC, GDGM, GPGM

1. Introduction

Companion diagnostics (CDx) have emerged as indispensable tools in precision medicine, particularly in guiding the effective use of targeted therapies for diseases. For a qualitative assay (e.g. gene mutation), the cutoff is typically the limit of detection (LOD) of the assay, its decision is relatively straightforward. For a quantitative assay or a semi-quantitative assay, determining the optimal cutoff for a CDx is complex. This cutoff categorizes patients into two groups based on their likelihood of responding to the drug. The choice of cutoff can significantly impact the clinical efficacy of the drug, the sample size requirements for clinical trials, and other vital parameters such as study

power and submission strategy. For example, if a cutoff is set too low, non-responders will be enrolled in the trial, the clinical efficacy will be 'diluted' and thus the trial requires a larger sample size to provide sufficient power. On the contrary, if the cutoff is set too high, some potential responders will miss the treatment opportunity and result in a reduced market size for the drug.

Although the CDx cutoff plays such an important role in the drug development and clinical trial strategies, it is challenging in practice since the decision-making process has to balance analytical uncertainty (demands more data and more studies) against drug development urgency (demands quick decisions with fewer studies). On the analytical side, the decision of the cutoff is based on phase 1 or phase 2 clinical trials where the sample size limited and the study design is not suited for well-structured analysis. Therefore, the resulted determination is inevitably highly 'uncertain' and 'risky'. In addition, the decision must also take into considerations of clinical factors such as prevalence, likelihood of future trial success, and even the analytical performance of the biomarker assay. While it is tempting to wait for more data so that a better-informed decision can be achieved, it is important to emphasize that any data used to determine the cutoff is considered a 'training data', the pivotal analysis has to exclude the data that were used in the cutoff decision. Therefore, from the clinical evidence point of view, it is desirable to determine the cutoff as early as possible so that the next trial serves as a 'validation' of the CDx. In addition, generally the next trial starts as soon as (or even before) the current trial finishes, it is thus time-sensitive that the decision needs to be made as quickly as possible.

This research investigates statistical methods and the decision strategy for determining the optimal cutoff for a CDx. In reality, the biomarker assay used in the

trials could be a clinical trial assay (CTA) which is not market ready. This research focuses on the cutoff decision; therefore, no distinction is made between CTA and CDx.

Receiver operating characteristic (ROC) curve analysis is often applied to measure the diagnostic accuracy of a biomarker. It is a plot of the true positive rate (TPR) on the Y-axis against the false positive rate (FPR) on the X-axis, at various cutoff value of a biomarker or a composite score of multiple biomarkers. The area under the ROC curve (AUC) provides an overview of the diagnostic accuracy of a biomarker, a biomarker with a bigger AUC suggests that it is better associated with the outcome. As a performance profile (TPR, FPR) at various cutoffs, a ROC curve is commonly used to determine the optimal cutoff value for a diagnostic product. For example, to develop a PD-L1 immunohistochemical (IHC) assay using the 22C3 anti–PD-L1 murine monoclonal antibody on the Dako platform as a possible companion diagnostic for pembrolizumab in patients with non–small cell lung cancer, the optimal scoring method and the cutoff score for predicting response to pembrolizumab with the clinical trial assay were identified by comparison using ROC curve analysis and were evaluated using the Youden index based on KEYNOTE-001.Youden Index maximizes the difference between true positive rate and false positive rate over all possible cutoff values. In addition to the Youden Index, several other methods have been proposed to determine the optimal cutoff point from ROC curves. The method by Perkins and Schisterman defines the optimal cutoff as the point that minimizes the Euclidean distance between the point on the ROC curve and the (0, 1) point. Miller & Siegmund's method maximizes the value of Chi-square statistic derived from the cross-tabulations of true disease status versus the classification results. Liu determines optimal cutoff by maximizing the product of sensitivity and specificity (concordance probability method). Unal proposed an alternative method called index of union (IU) that defines the optimal

cut-point value as the value whose sensitivity and specificity are the closest to the value of the area under the ROC curve and the absolute value of the difference between the sensitivity and specificity values is minimum. Hua & Tian proposed an adjusted Youden index method by penalizing misclassification rates of different orders.

It is worth noting that the method identifies the point closest to perfect point (0, 1) on the ROC curve involves a quadratic term and its clinical meaning is unknown. The IU method, Youden index and the concordance probability try to minimize the misclassification rate. All of these methods assume that the costs/risks of false positives and false negatives are equal, which are not true in most clinical situations. Unal recommended that researchers select the cutoff that is most clinically relevant. For example, the cutoff can be the point that provides the maximal TPR (FPR) while satisfying a desired FPR (TPR).

One major limitation of the methods based on the ROC curve is their insufficient consideration of the clinical risks. To address this limitation, this paper introduces a novel method for the optimal cutoff selection based on clinical trial Go/No-Go decision-making rule. Evidence based Go/No-Go decision making becomes essential step in the early phases clinical trials to accelerate development of promising compounds and stop development of inferior compounds early. Over the past decades, quantitative tools for decision making have been proposed. Lalonde et al. proposed a criterion based on the comparison of an asymmetric confidence interval to target values which provides a framework with 3 possible outcomes, namely, Go, Consider, and No-Go (3ODM: 3 Outcome Decision Making). Frewer et al. evaluate the features of the criteria of Lalonde et al. in further details based on a frequentist perspective and Pulkstenis et al. introduced the Bayesian approach of the same criterion. Walley et al. define a threshold of interest and derives a posterior distribution on the treatment effect given observed

data. A Go decision will be issued if the probability is sufficiently high for the efficacy exceeding this threshold.

Building on the 3ODM framework, a novel cutoff determination method, named Go decision guided method (GDGM), is proposed for determining the optimal cutoff for biomarkers in CDx development. The GDGM integrates the 3ODM framework with the cutoff determination process. It defines the optimal cutoff as the cutoff that maximizes the marker prevalence at which the decision is Go. This approach guarantees high enrolment rates (i.e. high prevalence for biomarker positive patients) while ensuring enough clinical efficacy to make a Go decision. In the GDGM approach, the decision-making relies on sample size. As an extension to GDGM, a Bayesian-based approach, termed the Go Probability Guided Method (GPGM), makes Go/No-Go decisions based on prediction using a fixed number of enrolled patients.

The remainder of the paper is organized as follows. Section 2 introduces the ROC method, the GDGM method, and the GPGM method for cutoff determination. Using IHC assay as an example biomarker, Section 3 describes the simulation methods for single-arm trials. Section 4 presents simulation settings and results, comparing the performance of the three methods. Section 5 discusses the limitations of GDGM and considers GPGM as an alternative method and discusses the potential directions for future research.

2. Methods for Cutoff Determination

2.1. ROC Method

The ROC curve is a graphical representation of the performance of a binary classification model across various cutoffs. It illustrates the diagnostic accuracy of the

classifier by plotting the TPR against the FPR at different classification cutoffs. From a clinical trial (simulated or real), let n_0 denote the number of non-responders and n_1 denote the number of responders. For each cutoff c, let n_0^0 denote the number of non-responders whose biomarker values are smaller than cutoff c (correctly classified as non-responders), and let n_0^1 denote the number of non-responders whose biomarker values are greater than cutoff c (incorrectly classified as responders). Similarly, let n_1^1 denote the number of responders whose biomarker values are greater than cutoff c (correctly classified as responders), and let n_1^0 denote the number of responders whose biomarker values are smaller than cutoff c (incorrectly classified as non-responders). FPR is, among those non-responders, the number of cases whose biomarker value is greater than the cutoff c.

$$FPR(c) = \frac{n_0^1}{n_0}$$

TPR is, among those responders, the number of cases whose biomarker value is greater than the cutoff c.

$$TPR(c) = \frac{n_1^1}{n_1}$$

The most commonly used ROC method (hereafter referred to as 'the ROC method') calculates the Euclidean distance between (0,1) and each point on the ROC curve. The Euclidean distance for a given cutoff c, $ED(c)$, is calculated as

$$ED(c) = \sqrt{(FPR(c) - 0)^2 + (TPR(c) - 1)^2}$$

The ROC method chooses the optimal cutoff that minimizes the Euclidean distance, i.e.,

$$Min_c\{ED(c)\}$$

Figure 1 provides an example ROC curve. The point (0, 1) is marked in red and the point closest to (0,1) is marked in blue. The grey dashed line represents the minimum value of $ED(c)$ across all cutoffs.

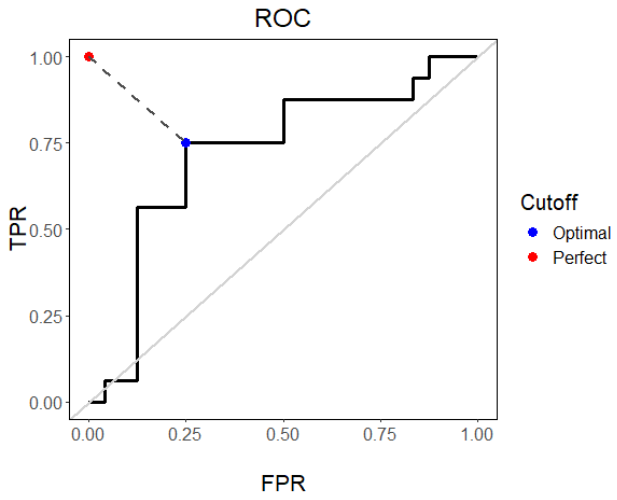


Figure 1. ROC Curve and the Optimal Cutoff. The red dot marks the perfect point, the blue dot indicates the selected cutoff, and grey dashed line represents the minimum distance between the ROC curve and the perfect point.

It is important to note that this method does not consider clinical goals (e.g. a minimum ORR that is clinically meaningful), nor does it consider clinical risk (e.g. whether a false positive or a false negative is clinically more harmful). It treats FPR and TPR in an equal manner and sets a goal function that has only mathematical rather than clinical interpretations. In terms of clinical decisions, one immediate problem with this method is that it can always identify an 'optimal cutoff' even if the treatment has low clinical efficacy or everyone has high likelihood of responding. This issue is not unique to this Euclidean-distance based ROC method; it is a limitation shared by the other ROC-based methods as well.

2.2. Go Decision Guided Method (GDGM)

GDGM is a cutoff selection method based on the 3ODM decision making framework. The 3ODM method was proposed by Frewer et al for clinical decision making. This method is built on multiple levels of targeted efficacy, extending the traditional Go/No-Go decision rule by introducing a Consider zone. This intermediate zone allows for smaller studies in early development to incorporate additional external or internal information to reach a definitive Go or No-Go decision. The criteria for these decisions are based on two pre-specified clinically justified parameters:

- Target value (TV), denoted by θ_{TV}, which is the desired level of performance.
- Lower reference value (LRV), denoted by θ_{LRV}, which is the smallest clinically meaningful treatment effect required for the development of a drug.

The decision in 3ODM is based on dual criteria related to the significance and relevance of the treatment effect. Let τ_L and τ_U represent the confidence level associated with these criteria.

- Significance Criterion requires $\tau_L\%$ confidence that the true treatment effect is larger than θ_{LRV}.
- Relevance Criterion requires $\tau_U\%$ confidence that the true treatment effect is not less than θ_{TV}.

A common choice for τ_L and τ_U are 0.8 and 0.9, respectively. The dual criterion can be assessed by confidence intervals (CI) using approaches such as Clopper Pearson method or Wald method from the frequentist perspective. The Significance Criterion is met if the lower bound of the one-sided $100 * \tau_L\%$ CI, denoted by $PCT_{1-\tau_L}$, is greater than θ_{LRV}. The Relevance Criterion is met if the upper bound of the one-sided $100 * \tau_U\%$

CI, denoted by PCT_{τ_U}, is not less than θ_{TV}. The dual criterion quantifies the strength of evidence from the observed results. This 3ODM framework by Frewer is as follows:

- A **Go** decision is made if the Significance Criterion and the Relevance Criterion are both met.
- A **No-Go** decision is made if the Relevance Criterion is not met.
- Otherwise, a **Consider** decision is made.

If the result falls within the No-Go or Go zones, the decision is straightforward. If the decision is Consider, additional guidance should be required to determine the next steps. This guidance might be based on a secondary metric or by comparing it with competitor data.

As an example, Figure 2 from Frewer et al illustrates the 3ODM decision rules.

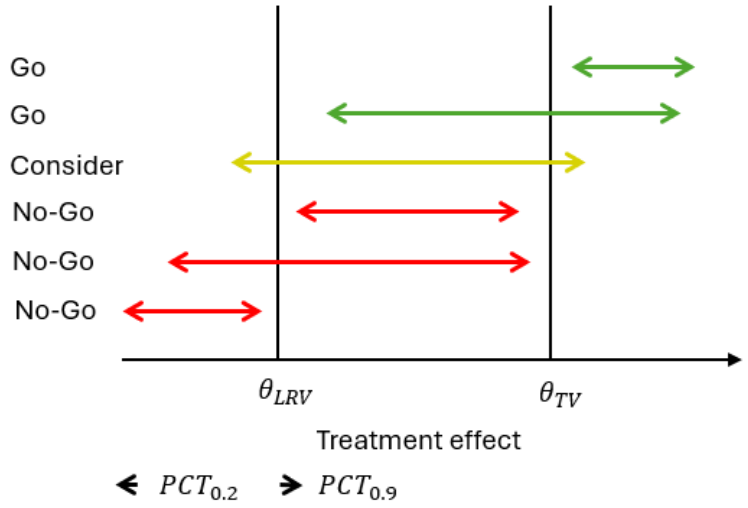


Figure 2. 3ODM Decision Rule . θ_{LRV} denotes the lower reference value, and θ_{TV} denotes the target value. A Go decision if $PCT_{0.2} > \theta_{LRV}$ and $PCT_{0.9} > \theta_{TV}$. A Consider decision if $PCT_{0.2} \leq \theta_{LRV}$ and $PCT_{0.9} > \theta_{TV}$. A Stop decision if $PCT_{0.9} \leq \theta_{TV}$.

Based on the 3ODM framework, the Go decision guided method (GDGM) works as follows for biomarker cutoff selection. For each candidate cutoff c, the clinical decision is evaluated under the framework. The optimal cutoff c is the one that maximizes the prevalence (i.e., the lowest cutoff) when the decision is Go. The condition of Go decision ensures the optimal cutoff achieves enough clinical efficacy. When the decision is Consider or No-Go, the clinical efficacy is not supported by the data, and thus the cutoff would not be selected. The cutoff determination principle of GDGM is given below.

$$Max_c\{Prevalence(c)\ |PCT_{1-\tau_L}(c) > \theta_{LRV}\ and\ PCT_{\tau_U}(c) > \theta_{TV}\}$$

2.3. Go Probability Guided Method (GPGM)

One issue of the GDGM method is that the selected patient number decreases as the cutoff increases. This leads to an increase in the width of the CI, and thus a decrease in reliability in the cutoff decision. This limitation is illustrated through simulations in Section 4. The Go Probability Guided Method (GPGM) makes decision based on a fixed number of enrolled patients (later this fixed number is annotated as N).

The GPGM method works in the following steps:

(1) For a pre-determined number of patients to be enrolled, N, calculate the minimal number of responders, x_{Go}, that leads to a Go decision. Based on the 3ODM framework, x_{Go} represents the minimum number of responders (out of N patients enrolled) for which the Significance Criterion and Relevance Criterion are both met. That is, the $PCT_{0.2}$ is greater than θ_{LRV} and the $PCT_{0.9}$ is greater than θ_{TV}.

(2) For each potential cutoff, assume that there are K patients with marker levels greater than or equal to this cutoff, and assume that there are R responders among these K patients. The ORR is R/K for this cutoff.

(3) Assume the trial continues and now only enrolls patients with marker level greater than or equal to this cutoff until a total of N patients are enrolled (i.e. another $N-K$ marker-positive patients are to be enrolled), calculate the likelihood of Go (i.e. the probability of observing $(x_{Go}-R)$ or more responders in the $(N-K)$ patients to be enrolled).

(4) Choose the lowest cutoff such that the Go probability in Step 3) is above a pre-determined threshold (e.g. 70%).

GPGM uses a Bayesian approach in Step 3 to determine the predictive probability of achieving the acceptance criteria when/if the target number of patients are enrolled. The probability is based on the posterior distribution of the observed efficacy in patients with marker value above the cutoff.

The predictive probability is calculated as follows. Assume that ORR follows a Beta distribution prior

$$ORR \sim Beta(a_0, b_0)$$

Given that R responders have been observed in K enrolled patients, the posterior for ORR follows a Beta distribution

$$ORR|R, K \sim Beta(a_0 + R, b_0 + K - R)$$

The number of future responders x in the remaining $N-K$ patients follow a Beta-Binomial distribution

$$x|R,K \sim Beta_Binomial(N-K\ , a_0+R, b_0+K-R)$$

Then, the predictive probability $P(Go)$ is the probability that $x_{total} = R + x$ is greater or equal to x_{Go}.

$$P(Go) = P(x_{total} \geq x_{Go}|R\ responders\ out\ of\ K) \quad (1)$$

$$= P(x \geq x_{Go} - R\ out\ of\ N-K\ |R\ out\ of\ K)$$

$$= \sum_{x=x_{Go}-R}^{N-K} \binom{N-K}{x} \frac{B(a_0+R+x, b_0+K-R+N-K-x)}{B(a_0+R, b_0+K-R)}$$

$$= \sum_{x=x_{Go}-R}^{N-K} \binom{N-K}{x} \frac{B(a_0+R+x, b_0+N-R-x)}{B(a_0+R, b_0+K-R)}$$

Where

- x represents the number of responders in the assumed patient enrollment.
- $B(.,.)$ is the beta function, defined as $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.
- A common choice of α and β are $a_0 = 0.5$, and $b_0 = 0.5$.

At the last step (Step 4), GPGM selects the lowest cutoff at which, if the trial continues to enroll a pre-specified total number of patients with marker value no less than this cutoff, the predictive probability of making a Go decision is greater than a certain threshold (e.g. 70%), i.e.,

$$Max_c\{Prevalence(c)\ |P(Go) > threshold\}$$

In comparison, GDGM method makes decision based on the observed data, the GPGM method calculates the likelihood of the "future" success.

3. Simulation Methods

3.1. Biomarker Data

The biomarker of the simulation is assumed to be the membranous staining percentage from an IHC assay. IHC is a laboratory technique that uses antibodies to detect specific proteins (antigens) in tissue samples. In practice, the staining percentage is either captured as continuous data or grouped into discrete bin levels. It should be noted that this research is based on an IHC assay, the same idea can be applied to any quantitative or semi-quantitative biomarker.

In the simulations, the biomarker data are considered under both continuous and binned settings. For continuous biomarkers, normal distribution is considered to reflect the distributional characteristics of the typical biomarker data. For binned biomarker data, the distribution – bin levels along with their prevalences – is constructed to mimic a real-life setting. In addition, a real biomarker dataset is also included.

3.2. Relationship Between Biomarker and Response

Denote the ordered biomarker data (the membranous staining percentage of the IHC assay) by $\mathbf{x} = (x_1, \dots, x_N)$, where N is the sample size. The probability of response $r(x_i)$ is assumed to be a monotone increasing function of the biomarker value with lower and upper response bounds denoted by r_{min} and r_{max}, respectively. The following biomarker-response relationships are considered in the simulations.

Linear Relationship

$$r(x_i) = r_{min} + (r_{max} - r_{min}) \times \frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \quad (2)$$

Stepwise relationship

For the binned biomarker, a monotone stepwise relationship is assumed. Suppose that the biomarker values are divided into k bins $[x_0, x_1), \ldots, [x_{k-1}, x_k]$. Let p_j denote the prevalence of subjects with marker expression level in bin $[x_{j-1}, x_j)$, and let r_j denote the response rate for bin $[x_{j-1}, x_j)$, with $r_1 \leq \cdots \leq r_k$. Note that r_j can also be interpreted as the probability of response for each of the subjects in bin $[x_{j-1}, x_j)$. The biomarker-response function is defined as

$$r(x_i) = \begin{cases} r_1, x_i \in [x_0, x_1) \\ \quad \ldots \\ r_k, x_i \in [x_{k-1}, x_k] \end{cases} \tag{3}$$

In the simulation, with assumed biomarker data and biomarker-response relationship, the response outcomes are generated from Binomial distributions. For continuous biomarkers, the response status for a subject with biomarker value of x_i is generated by Binomial$(1, r(x_i))$ (which is Bernoulli with probability $r(x_i)$). For the binned biomarker, the number of responders for each bin is generated from a Binomial distribution with size of $p_i N$ and probability of r_i.

3.3. Theoretical Cutoff by Method

With the ordered biomarker values $\mathbf{x} = (x_1, \ldots, x_N)$, candidate cutoffs are evaluated by a grid search approach over the range from the 10th to the 90th percentiles of the biomarker data. The reason for searching within [10%, 90%] of the data range is that (i) if the CDx selects >90% of the patients, then practically the CDx is not needed; (ii) when the cutoff is so high that only <10% of the patients are left, the statistical estimates (point estimate and standard error) become unreliable. For a cutoff $c = x_i$, let

$$Selected(c) = \{j: x_j \geq c\}$$

denote the set of patients selected by the cutoff c and let K be the number of selected patients. The theoretical ORR is calculated as the average of the response rates (probabilities) across the selected patients. The theoretical ORR is given by

$$ORR(c) = \frac{1}{K} \sum_{j \in Selected(c)} r(x_j)$$

Where $r(x_j)$ is defined by Equation (2) or (3). When the stepwise relationship in Equation (3) is used, all subjects within the same bin share the same response probability. For example, if the kth bin is selected, all $p_k N$ subjects in this bin have response probability r_k, that is, $r(x_j) = r_k$ for $x_j \in [x_{k-1}, x_k]$. In this case, the ORR is equal to a weighted average of the bin-specific response probabilities.

Theoretical Cutoff for GDGM

Using Wald method, the theoretical lower bound of the $100 * \tau_L\%$ one-sided CI, denoted by $PCT_{1-\tau_L}$, is given by

$$PCT_{1-\tau_L}(\mathrm{c}) = \max\left\{ ORR(c) - z_{\tau_L} \sqrt{\frac{ORR(c)(1-ORR(c))}{N(c)}}, 0\% \right\}$$

Similarly, the upper bound of the $100 * \tau_U\%$ one-sided CI, denoted by PCT_{τ_U}, is

$$PCT_{\tau_U}(\mathrm{c}) = \min\left\{ 100\%, ORR(c) + z_{\tau_U} \sqrt{\frac{ORR(c)(1-ORR(c))}{N(c)}} \right\}$$

Note that $PCT_{1-\tau_L}(c)$ and $PCT_{\tau_U}(c)$ can also be calculated based on other methods (e.g.

Clopper-Pearson method, Wilson score method). Based on the theoretical CI bounds, the theoretical cutoff for GDGM is the smallest cutoff such that $PCT_{1-\tau_L}(c) > \theta_{LRV}$ and $PCT_{\tau_U}(c) > \theta_{TV}$.

Theoretical Cutoff for GPGM

For GPGM, with a pre-determined sample size N, x_{Go} can be calculated. A Jefferys prior of a Beta distribution is assumed for ORR. For each cutoff c, the number of subjects above this cutoff, K, as well as the theoretical number of responders, $R = \sum_{j \in Selected(c)} r(x_j)$, can be determined. The theoretical probability of reaching a Go decision upon the completion of enrolling N patients when already observing R responders out of K selected patients can be calculated for this cutoff. The theoretical cutoff for GPGM is the smallest cutoff such that $P(Go) > p_0$ (e.g. $p_0 = 0.7$).

Theoretical Cutoff for ROC

For ROC, the theoretical TPR and FPR are

$$TPR(c) = \frac{\sum_{j \in Selected(c)} r(x_j)}{\sum_{j \in \{1,\dots N\}} r(x_j)}$$

$$FPR(c) = \frac{\sum_{j \in Selected(c)} r(x_j)}{\sum_{j \in \{1,\dots N\}} r(x_j)}$$

The theoretical ROC cutoff is the one that achieves the minimum Euclidean distance to (0,1).

3.4. Calculation Procedures

For a cutoff c, the ORR is estimated by

$$\widehat{ORR}(\mathrm{c}) = \frac{m(c)}{N(c)}$$

where $m(c)$ is the number of responders above the cutoff c, and $N(c)$ is the number of subjects above the cutoff c.

Using Wald method, the confidence interval for ORR, $[100 * \tau_L\%, 100 * \tau_U\%]$, is given by

$$\widehat{PCT}_{1-\tau_L}(\mathrm{c}) = \max\left\{\widehat{ORR}(\mathrm{c}) - z_{\tau_L}\sqrt{\frac{\widehat{ORR}(c)(1-\widehat{ORR}(c))}{N(c)}}, 0\%\right\}$$

$$\widehat{PCT}_{\tau_U}(\mathrm{c}) = \min\left\{100\%, \widehat{ORR}(\mathrm{c}) + z_{\tau_U}\sqrt{\frac{\widehat{ORR}(c)(1-\widehat{ORR}(c))}{N(c)}}\right\}$$

For GDGM, by comparing $\widehat{PCT}_{1-\tau_L}(\mathrm{c})$ with θ_{LRV} and $\widehat{PCT}_{\tau_U}$ with θ_{TV}, this approach can categorize the clinical decision into three outcomes: Go, Consider, and No-Go according to 3ODM framework. The optimal cutoff is determined as the value that maximizes the prevalence of biomarker positive under the Go decisions.

For GPGM, with a pre-determined number of patients to be enrolled, N, the minimal number of responders that leads to a Go decision, x_{Go}, is also determined based on the Go criteria. For example, if the pre-determined sample size is $N = 40$, $x_{Go} = 16$. The predictive Go probability $P(Go)$ is computed according to Equation (1) when the prior is assumed to follow a Beta (0.5, 0.5) distribution. The optimal cutoff is selected as the value that maximizes the prevalence of biomarker positive when $P(Go) > 70\%$.

4. Simulation Study

The simulation study assumes that a single-arm trial enrolled $N = 40$ patients. LRV is θ_{LRV}=20% and TV is θ_{TV}=50%. The significance level is τ_L =0.8 and the relevance level is τ_U =0.9. The simulation framework can be extended to a two-arm trial, the conclusions remain the same as the single-arm trial. Some simulations are presented in the Appendix.

4.1. Simulation Scenarios

Three scenarios are considered. The details for the simulation scenarios are provided in Table 1, which includes the biomarker distribution and assumed biomarker-response relationships for each setting.

- Scenario 1 assumes that the biomarker values are continuous and follow a normal distribution. The response probability has a linear relationship with the biomarker value following Equation (2). Under this setting, a Go decision is supported for selected cutoffs. Specifically, the dual criteria $PCT_{0.2}(c) > \theta_{LRV}$ and $PCT_{0.9}(c) > \theta_{TV}$ are satisfied for cutoffs of 20 and above, and the Go probability exceeds 70% for the same range of cutoffs.
- Scenario 2 assumes binned biomarker data, and the biomarker-response relationship follows the step function in Equation (3). In this case, the dual criteria are met for cutoffs of 10 and above, while the Go probability exceeds 70% for cutoffs of 30 and above.
- Scenario 3 uses an empirical biomarker dataset in which expression values (membranous staining percentage from an IHC assay) are concentrated at the two ends, with fewer observations in the middle. The response is assumed to follow a linear relationship as specified in Equation (2). The dual criteria are

again satisfied for cutoffs of 10 and above, and the Go probability exceeds 70% for cutoffs of 30 and above.

Table 1. Simulation Settings.

	Scenario 1	**Scenario 2**	**Scenario 3**
Biomarker Distribution	Normal (mean=30, SD=10)	Bin (with prevalence) • [0, 1): 5% • [1, 5): 10% • [5, 10): 10% • [10, 30): 25% • [30, 50): 5% • [50, 70): 25% • [70, 100]: 20%	Empirical (Real Data)
Biomarker-Response Relationship	Linear Equation (2) $r_{min} = 5\%$ $r_{max} = 70\%$	Stepwise Equation (3) $(r_1, \dots, r_7) =$ (5%, 10%, 15%, 20%, 35%, 50%, 55%)	Linear Equation (2) $r_{min} = 5\%$ $r_{max} = 75\%$

Figure 3 shows the histograms of biomarker data and theoretical cutoffs. Following the theoretical calculations in Section 3.3, theoretical cutoffs for GDGM, GPGM and ROC are marked in red, green, and blue dotted lines. The LRV and TV values are shown by a dash-dot line and a dashed line, respectively. Among the three scenarios, the theoretical ROC cutoffs are higher than the theoretical cutoff by GDGM or GPGM.

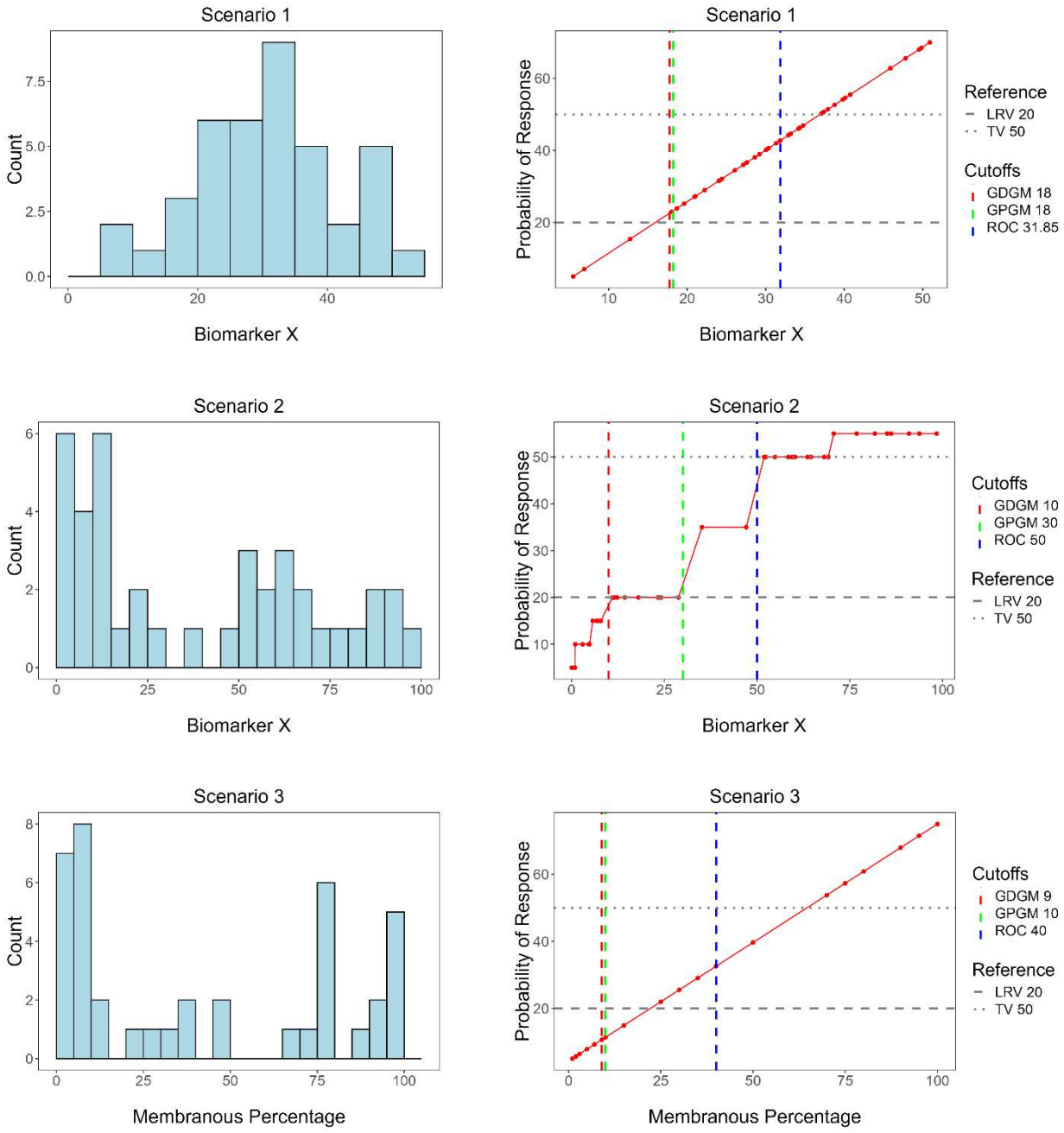


Figure 3. Biomarker Distributions and Theoretical Cutoffs. Left: Biomarker Distributions. Right: Theoretical Cutoffs. The theoretical cutoffs for GDGM, GPGM, and ROC are indicated by red, green, and blue dotted lines, respectively. The LRV and TV values are shown by a dash-dot line and a dashed line, respectively.

4.2. Simulation Results

Figure 4 below illustrates how GDGM, GPGM and ROC method selected the cutoff

from one simulation run.

In the GDGM plots (left panel), the blue dotted line, red solid line, and green dotted line represent $PCT_{0.2}$, the observed ORR, and $PCT_{0.9}$, respectively. The blue and green vertical solid lines represent the θ_{LRV} and θ_{TV}, respectively. The selected cutoff is represented by the black dashed horizontal line. It shows that GDGM selected reasonable cutoffs in all 3 Scenarios.

In the GPGM plots (middle panel), the green line represents $P(Go)$. The red vertical solid line is at 70%. The selected cutoff is represented by the black dashed horizontal line. It shows that GPGM selects a cutoff greater or equal to the GDGM cutoff in all Scenarios. This occurs because, in the GDGM plots, the upper confidence bound exceeds θ_{TV} due to large variability, whereas the corresponding predictive Go probability in the middle panels is not sufficiently high. Therefore, GPGM selects a higher cutoff to ensure a Go decision.

In the ROC plots (right panel), the red dot marks the perfect point, and the blue dot indicates the selected cutoff. Across all scenarios, ROC method identifies the midpoint of the biomarker range as the optimal cutoff for distinguishing responders from nonresponders. In Scenarios 1, where biomarker values range from 0 to 60, the optimal cutoff is approximately 30. In Scenario 2 and 3, where biomarker values range from 0 to 100, the optimal cutoffs are approximately 50.

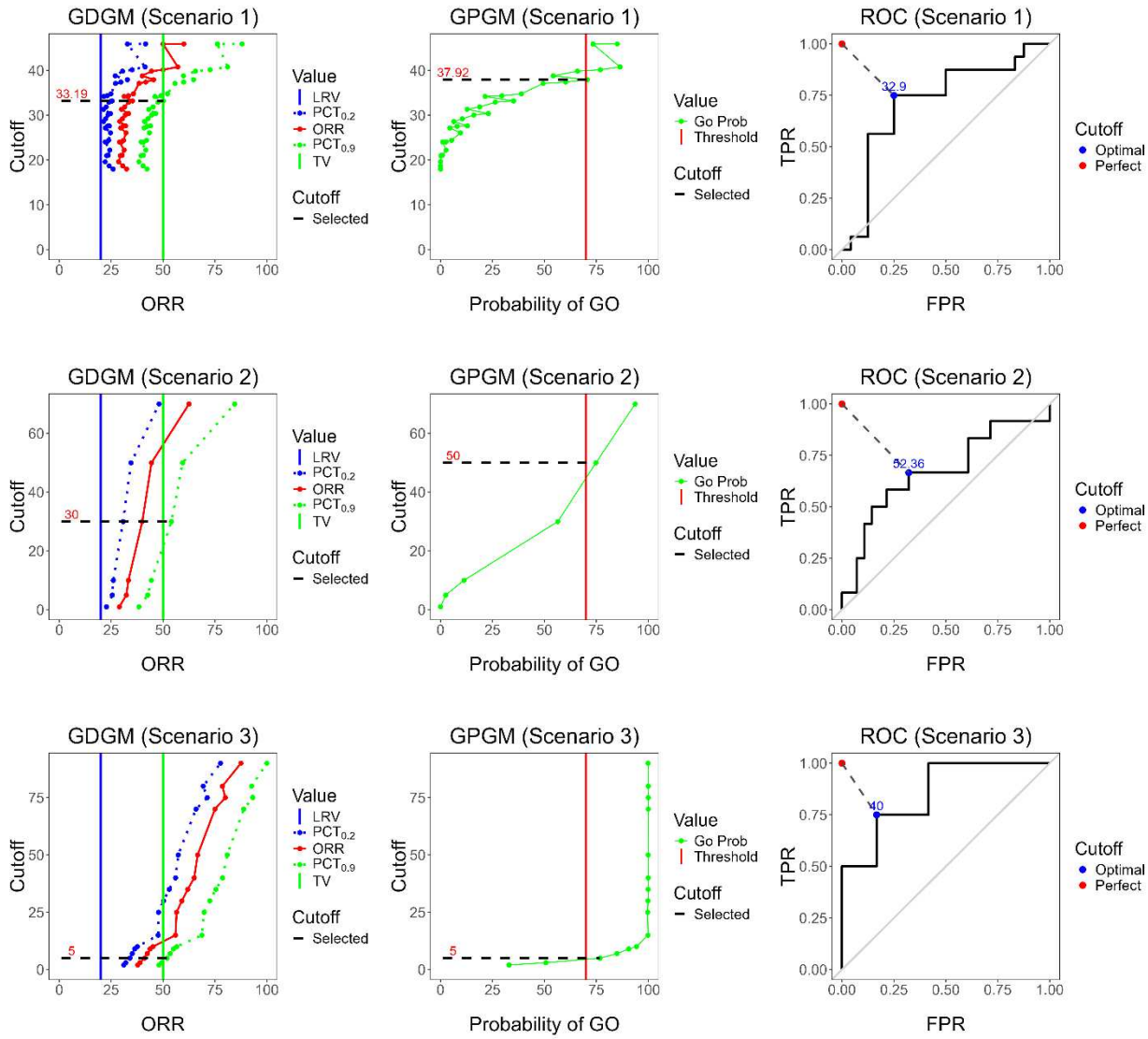


Figure 4. GDGM, GPGM ROC Cutoff for One-arm Trial (One Simulation). Left: GDGM results. Blue dotted, red solid, and green dotted curves represent $PCT_{0.2}$, observed ORR, and $PCT_{0.9}$, respectively. The vertical solid lines indicate θ_{LRV} (blue) and θ_{TV} (green). The selected cutoff is shown as a black dashed horizontal line. Middle: GPGM results. The green curve represents P(Go). The red vertical line denotes the 70% Go threshold. The selected cutoff is shown as a black dashed horizontal line. Right: ROC-based cutoff selection. The red dot marks the perfect point, the blue dot indicates the selected cutoff, and grey dashed line represents the minimum distance between ROC curve and the perfect point.

Figure 5 shows the selected cutoffs by the GDGM, GPGM and ROC methods from 1000 simulation runs. The left panel presents the distributions of selected cutoffs for each method. The right panel shows the corresponding cumulative probability (CP), in which the 80th percentile is indicated. This 80th percentile is also marked on the left panel plots by a vertical dotted line for GDGM (red), GPGM (green) and ROC (blue). This 80th cumulative probability line indicates cutoff above which 80% of the simulations results are Go. If this line intersects with the bars, it suggests that a Go decision was made in at least 80% of simulations at or below that cutoff. In other words, if the cumulative Go rate at a cutoff intersects with the 80% horizontal line, we have ≥80% confidence that this cutoff will lead to a Go decision. Conversely, if the bars do not intersect with this horizontal line, it indicates that none of the cutoffs resulted in Go with high confidence. For example, in Scenario 1, GDGM achieves a Go rate exceeding 80% at a cutoff of 21, whereas GPGM reaches an 80% Go rate at a cutoff of 24. In contrast, the ROC method selects a higher cutoff of 34 to achieve the same 80% confidence level.

With the random variation across simulation runs, each of which either results in a Go decision with an optimal cutoff or fails to meet the Go criteria with no cutoff chosen, the 80th percentile represents a 'summary statistics' of the selected cutoffs. Study team could discuss and align on the percentile threshold used to determine the cutoff. For example, the 90th percentile can be used if a more conservative (high) cutoff is preferred. If one cutoff satisfies the Go decision, a higher cutoff would also ensure a Go decision due to higher ORR, but it is less aggressive in terms of prevalence (i.e. prevalence is lower).

Several observations can be made. First, Scenario 1 considers normally distributed biomarker data with a linear biomarker-response relationship. Consistent

with the pattern observed for the theoretical cutoffs, the 80th percentile by ROC is higher than those by GDGM and GPGM. This indicates that GDGM and GPGM tend to favor lower cutoffs, effectively maximizing the prevalence of positive biomarkers while maintaining sufficient clinical efficacy. In contrast, the ROC method favors higher clinical efficacy by choosing a higher cutoff that results in lower prevalence.

Scenario 2 examines the binned biomarker setting with a monotone increasing stepwise biomarker-response relationship. As in the theoretical results shown in Figure 3, the 80th percentile selected by GPGM is higher than that selected by GDGM. Although the theoretical cutoff selected by GPGM is lower than that selected by the ROC method, the 80th percentiles for these two methods (indicated by the green line and blue line) coincide at 50. This result is reasonable because the ROC method determines the optimal cutoff solely based on the observed data, whereas GPGM incorporates predictive information, introducing additional uncertainty. Consequently, even when the theoretically selected cutoffs differ, accounting for variability can lead to the same 80th percentile. This finding also suggests that despite being a commonly used approach, the ROC method performs no better than GPGM in this setting.

Moreover, the behavior of the ROC method reflects its design objective of maximizing separation between responders and non-responders. Given that relatively few patients fall into the middle bin [30,50), ROC tends to select a cutoff near this region. In contrast, GDGM and GPGM are designed to include as many patients as possible while still maintaining adequate efficacy, leading to different cutoff selection distribution.

Scenario 3 represents an empirical biomarker distribution in which marker values are concentrated at the two extremes but less frequent in the middle. The 80th percentile of the cutoffs selected by both GDGM and GPGM is 15, whereas the ROC-

based method yields a cutoff of 75. At the cutoff of 75, the theoretical ORR is 67%, exceeding the target θ_{TV}, but only 37.5% of patients (15/40) are classified as biomarker positive. In contrast, at the cutoff of 15, the theoretical ORR is 45.83%, with a larger proportion of patients 62.5% (25/40) classified as biomarker positive. The overall conclusion remains consistent that GDGM and GPGM achieve a better balance between clinical efficacy and biomarker prevalence than the ROC-based method.

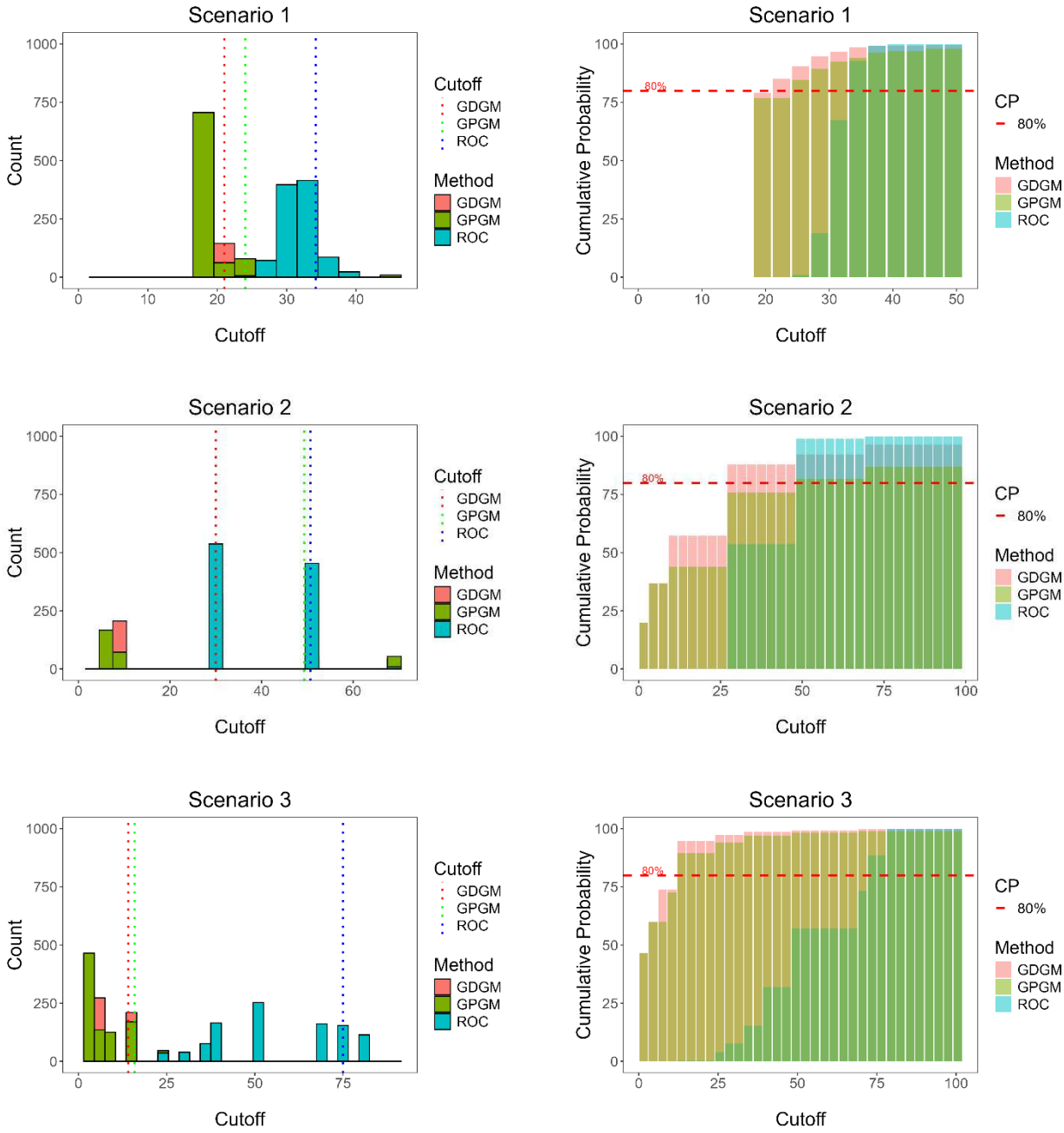


Figure 5. Cutoff Determination by GDGM, GPGM, and ROC (1000 Simulations). Left: distributions of selected cutoffs. The 80th percentile is marked by a vertical dotted line for GDGM (red), GPGM (green), and ROC (blue). Right: the cumulative probability (CP). The 80th percentile is indicated by the red horizontal line.

5. Discussion

This paper proposes a novel method, GDGM, for optimal biomarker cutoff selection based on clinical decision rules. Compared to the ROC method, GDGM ensures that the selected cutoff achieves sufficient clinical efficacy (e.g. meet Go criteria in clinical decision-making framework) while maximizing the number of patients who can benefit from the treatment (biomarker prevalence). The proposed optimization principle can be applied to various clinical trial decision rules, with the three-outcome decision-making (3ODM) framework being one example.

The reliability for Go decision under GDGM depends on the sample size in the selected biomarker positive subgroup. As an extension of GDGM, GPGM makes decisions and establishes the optimal cutoff based on the predictive Go probability for a fixed sample size. The performance of GPGM depends on the assumptions for the Bayesian prior Beta (0.5, 0.5), also known as Jeffreys prior, was used in the simulation. Though non-informative, its U-shaped distribution places more weights near 0 and 1 than around the mean. Careful consideration of the prior distribution is advised when applying GPGM in practice.

Multiple settings were considered in the simulations. In terms of the biomarker data, the simulations considered ordinal marker with discrete bin levels and quantitative biomarker with continuous expression values (distributions derived from normal distributions). Additional biomarker distributions could be considered in future study. In

terms of the biomarker-response relationship, the simulations in this research assumed that the response rate increases monotonically with the biomarker value. Specifically, when the biomarker is quantitative (Scenario 1 and 3), the simulations assumed a linear relationship. The appendix provides more simulation results when the relationship is non-linear. In practice, the relationship between biomarker levels and treatment response may take different forms (e.g. linear or non-linear), such as a non-decreasing function that the response rate is flat after a biomarker threshold. In terms of the study design, the simulations focused on single-arm clinical trials. The analysis for a two-arm design becomes more complicated. The first challenge is determining if the biomarker is indeed predictive, which requires a large randomized two-arm design to test the treatment-marker interaction effect, and this is practically unfeasible. Second, a predictive biomarker can be predictive but not prognostic, or can be both predictive and prognostic. This knowledge would inform how the response in the control arm should be simulated. Third, with the number of responders in the control arm becoming a random variable at each cutoff, the determination of x_{Go} and the Go probability for GPGM becomes more complicated. Application of GPGM in the two-arm setting will be of interest for future research.

The focus of the work is on the early oncology clinical endpoint ORR. In clinical trial development, the standard primary endpoint for pivotal study is longer-term clinical outcomes such as progression-free survival (PFS) and overall survival (OS). The potential disconnect between early and late endpoints should be carefully considered in future studies. In addition, although oncology trials are used as the motivating example here, extending the methodology to other therapeutic areas would be of interest.

Disclosure Statement

No potential competing interest was reported by the author(s).

Data Availability Statement

The datasets will be available from the corresponding author on reasonable request.

Ethics Statement

This study did not involve human participants or animal subjects. Therefore, ethical approval and informed consent were not required.

Funding Statement

No funding was received for this study.

References

Dolled-Filhart, M., Roach, C., Toland, G., Stanforth, D., Jansson, M., Lubiniecki, G. M., Ponto, G., & Emancipator, K. (2016). Development of a Companion Diagnostic for Pembrolizumab in Non-Small Cell Lung Cancer Using Immunohistochemistry for Programmed Death Ligand-1. *Archives of pathology & laboratory medicine*, *140*(11), 1243–1249. https://doi.org/10.5858/arpa.2015-0542-OA

Youden, W. J. *Index for rating diagnostic tests.* Cancer. https://doi.org/10.1002/1097-0142(1950)3:1%3C32::aid-cncr2820030106%3E3.0.co;2-3

Fluss, R., Faraggi, D., & Reiser, B. (2005). *Estimation of the Youden index and its associated cutoff point.* Biometrical Journal. https://doi.org/10.1002/bimj.200410135

Perkins, N. J., & Schisterman, F. E. (2006). *The inconsistency of "optimal" cut-points using two ROC based criteria.* American Journal of Epidemiology. https://doi.org/10.1093/aje/kwj063

Miller, R., & Siegmund, D. (1982). *Maximally selected chi square statistics.* Biometrics. Journal of the Biometric Society. https://doi.org/10.2307/2529881

Liu, X. (2012). *Classification accuracy and cut point selection.* Statistics in Medicine. https://doi.org/10.1002/sim.4509

Unal, I. (2017). *Defining an optimal cut-point value in ROC analysis: an alternative approach.* Computational and mathematical methods in medicine. https://doi.org/10.1155/2017/3762651

Hua, J., & Tian, L. (2019). *A comprehensive and comparative review of optimal cut-points selection methods for diseases with multiple ordinal stages.* Journal of Biopharmaceutical Statistics. https://doi.org/10.1080/10543406.2019.1632876

Lalonde, R. L., Kowalski, K. G., Hutmacher, M. M., Ewy, W., Nichols, D. J., Milligan, P. A., Corrigan, B. W., Lockwood, P. A., Marshall, S. A., Benincosa, L. J., Tensfeldt, T. G., Parivar, K., Amantea, M., Glue, P., Koide, H., & Miller, R. (2007). *Model-based drug development*. Clinical pharmacology and therapeutics, 82(1), 21–32. https://doi.org/10.1038/sj.clpt.6100235

Frewer, P., Mitchell, P., Watkins, C., & Matcham, J. (2016). *Decision-making in early clinicaldrug development*. Pharmaceutical Statistics. https://doi.org/10.1002/pst.1746

Pulkstenis, E., Patra, K., & Zhang, J. (2017). *A Bayesian paradigm for decision-making in proof-of-concept trials*. Journal of Biopharmaceutical Statistics, 27(3), 442–456. https://doi.org/10.1080/10543406.2017.1289947

Walley, R. J., Smith, C. L., Gale, J. D., and Woodward, P. (2015), *Advantages of a wholly Bayesian approach to assessing efficacy in early drug development: a case study*. Pharmaceut. Statist., 14, 205–215. https://doi.org/10.1002/pst.1675

Appendix

A1. Additional Simulation Results from Single-Arm Trial

In addition to simulation scenarios based on linear and stepwise biomarker–response relationships, results for a nonlinear relationship, specifically a quadratic (square) function, are also presented. In this scenario, the biomarker distribution is the same as in Scenario 1 of the main text, while the biomarker–response relationship follows a square function.

$$r(x_i) = r_{min} + (r_{max} - r_{min}) \times \left(\frac{x_i - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})} \right)^2$$

where $r_{min} = 5\%$ and $r_{max} = 85\%$.

Figure A1 shows the histograms of biomarker data and theoretical cutoffs.

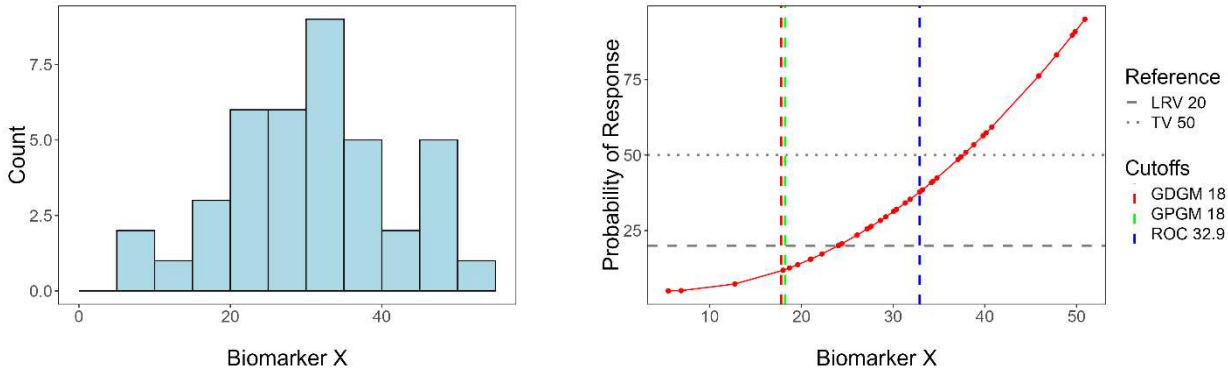


Figure A1. Biomarker Distributions and Theoretical Cutoffs. Left: Biomarker Distributions. Right: Theoretical Cutoffs. The theoretical cutoffs for GDGM, GPGM, and ROC are indicated by red, green, and blue dotted lines, respectively. The LRV and TV values are shown by a dash-dot line and a dashed line, respectively.

Figure A2 below illustrates how GDGM, GPGM and ROC method selected the cutoff from one simulation run.

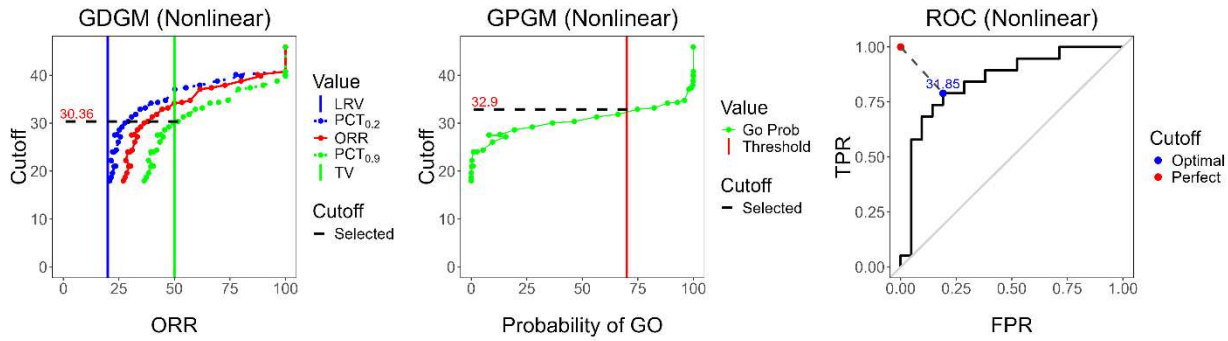


Figure A2. GDGM, GPGM ROC Cutoff for One-arm Trial (One Simulation). Left: GDGM results. Blue dotted, red solid, and green dotted curves represent $PCT_{0.2}$, observed ORR, and $PCT_{0.9}$, respectively. The vertical solid lines indicate θ_{LRV} (blue) and θ_{TV} (green). The selected cutoff is shown as a black dashed horizontal line. Middle: GPGM results. The green curve represents P(Go). The red vertical line denotes the 70% Go threshold. The selected cutoff is shown as a black dashed horizontal line. Right: ROC-based cutoff selection. The red dot marks the perfect point, the blue dot indicates the selected cutoff, and grey dashed line represents the minimum distance between ROC curve and the perfect point.

Figure A3 shows the distribution of selected cutoff by GDGM, GPGM, and ROC. The conclusion remains consistent, the GDGM method achieves a better balance between clinical efficacy and marker prevalence compared to the ROC-based method.

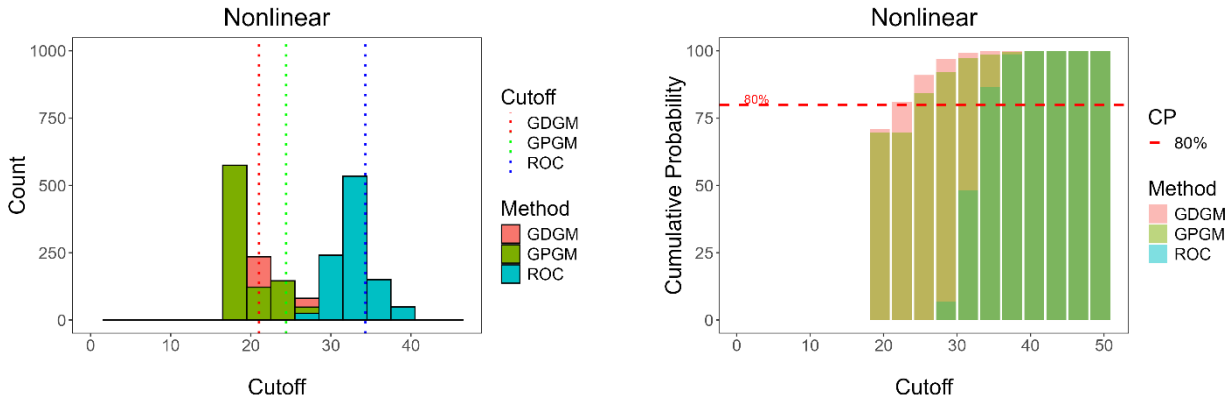


Figure A3. Cutoff Determination by GDGM, GPGM, and ROC (1000 Simulations). Left: distributions of selected cutoffs. The 80th percentile is marked by a vertical dotted line for GDGM (red), GPGM (green), and ROC (blue). Right: the

cumulative probability (CP), in which the 80th percentile is indicated by the red horizontal line.