[WIP] Best Practice for Communicating Probability of Success to Non-Statisticians

# Introduction

Communicating the probability of success (PoS) in clinical trials to non-statisticians (e.g. clinicians, project leads, senior management) is crucial for informed and impactful decision-making.

While PoS serves as a vital measure for use in drug development decision-making, its inherent statistical nature can pose a communication challenge. The complexities of probability, statistical models, and data interpretation may not be readily grasped by individuals without a strong statistical background. This gap in understanding can lead to misinterpretations, misaligned expectations, and ultimately, suboptimal decision-making. This can lead to misuse of resources and delayed drug development.

**💡 Helpful Tip:** When introducing PoS, explain that it's your team's best current estimate of whether a trial will meet its goals. Emphasize that it’s a key metric to help everyone understand the chances of success and collaboratively make informed decisions.

This document aims to provide best practices for statisticians and quantitative disciplines to effectively convey these statistical concepts in a clear and understandable manner, ensuring that stakeholders without a statistical background can grasp the importance and implications of PoS in the context of drug development. The document contains stand alone sections which can be reviewed independently of each other.

**Table 1: Definitions and Abbreviations**

**💡 Helpful Tip:** When presenting, reassure your non-statistician audience not to be daunted by technical terms. Provide a table as a handy reference for any unfamiliar terminology encountered during the discussion.

|  |  |
| --- | --- |
| **Term** | **Definition** |
| Conditional Probability | Likelihood of an event given specific conditions |
| Consider Region | Inconclusive results between MV and TV |
| Go Region | Results at or above TV |
| Minimum Value (MV) | The lowest acceptable outcome for success. |
| No Go Region | Results below MV |
| Probability of Success (PoS) | The likelihood of achieving a successful outcome |
| Statistical Power | The probability of detecting a true effect if it exists |
| Target Value (TV) | The desired optimal outcome |
| Unconditional Probability | Likelihood of achieving objectives across a range of scenarios |

# Using Simulation and Graphics to Communicate PoS

Simulations and visualizations are powerful tools for communicating statistical concepts like PoS to non-statisticians. These approaches help relate complex topics into insights that support understanding and decision-making across cross-functional teams.

By simulating trial outcomes under varying assumptions—such as different sample sizes, prior distributions, or treatment effect sizes—statisticians can demonstrate how key inputs affect study design and the associated PoS. Visualizations of these simulations not only make the results more interpretable but also help non-statisticians understand the implications of uncertainty in a more intuitive way. Throughout this guidance, different visualizations are proposed to support communication to non-statisticians.

# Defining Success

Success in drug development is complex, partly because it often spans multiple disciplines such as clinical, regulatory, operational and commercial, and partly because the concept of success evolves as a drug moves through the development process.

**💡 Helpful Tip:** Stress to your audience that "success" is multifaceted. Use an analogy: explain that a clinician might define success by patient improvement, a regulator by the availability of sufficient evidence for approval, and commercial teams by market viability. Frame your role as helping to *quantify* and *integrate* these diverse success factors (e.g., clinical efficacy, regulatory endorsement, market access viability, operational feasibility) into a cohesive PoS assessment, especially as the drug progresses and definitions of success evolve.

Traditionally, statisticians focussed on ensuring robust design for regulatory studies by utilising statistical techniques to design an appropriately powered study to detect a meaningful effect supporting regulatory approval.

More recently, there has been a greater focus on interim analyses and decision rules which will de-risk a study and provide early evidence to stop early (either for efficacy or futility), continue with greater confidence or continue with some adaptation(s) to the study.

Statisticians are increasingly involved at a strategic level, extending beyond individual clinical trials by considering the entire development program with different decision points and success criteria required at each step. By considering how these criteria and decisions depend on each other, statisticians play a crucial role in supporting a project or development team to define, quantify, and maximise success.

**Different types of success**

Different stakeholders will have different ideas of what success might entail.

To ensure alignment among stakeholders, it is essential to define success criteria clearly. These criteria may vary depending on the perspective of different stakeholders, such as clinical teams, regulatory bodies and investors. Key success criteria include:

* **Clinical Success**
  + **Efficacy**: Demonstrating a meaningful therapeutic effect, including both statistical significance and clinically relevant benefit.
  + **Safety**: Ensuring an acceptable benefit-risk profile.
* **Regulatory Success**
  + Approval Readiness: Meeting the requirements of regulatory agencies, including alignment with expectations for efficacy and safety.
* **Operational Success**
  + Clinical Operations: Achieving timely recruitment and data availability for interim or final analyses, especially when strategic or competitive timelines are involved.
  + Manufacturing: Ensuring the drug can be produced to specification.
* **Market Success**
  + Market Potential: Aligning with a Target Product Profile (TPP) that supports sales targets and market share.
  + Payer Value: Demonstrating clinical and economic value that supports payer expectations and reimbursement strategies.

**💡 Helpful Tip:** Explain these key criteria as a checklist. Emphasize that overall program success often requires ticking multiple boxes, not just achieving statistical significance on one endpoint.

These success criteria should be combined in a stepwise fashion, for example:

* **Phase 3 study success (Efficacy X Safety):**
  + Obtaining statistically significant and clinically relevant results that support the drug effectiveness in one or multiple studies, with a proportionate and acceptable safety profile in the context of the drug therapeutic benefits.
* **Regulatory success (Efficacy X Safety X Regulatory):**
  + Meeting phase 3 study success criteria and having drug approved by target regulatory agencies.
* **Market Success (Efficacy X Safety X Regulatory X Market):** 
  + Regulatory success whilst also meeting TPP targets.

Sometimes the different types of success are inherently linked and shouldn’t be assumed as independent success probabilities that can be multiplied together.

**💡 Helpful Tip:** You can illustrate how PoS can help evaluate trade-offs: "For instance, if we broaden inclusion criteria to accelerate Phase 3 recruitment (an operational win), this might impact the observed treatment effect or safety profile, and consequently, the PoS for regulatory approval”

Balancing diverse success objectives to design an optimal study or development program requires input from multiple stakeholders and a clear understanding of the interrelationships among key factors.

# Success of an Individual Study

At the individual study level, success is typically defined using one of the following approaches:

* Statistically significant difference favoring the test treatment over the reference treatment (i.e., rejection of the null hypothesis of no test treatment effect). This is the most commonly used definition of study success, particularly in Phase 3 studies.
  + A more stringent definition of success is sometimes applied, requiring not only statistical significance but also that the point estimate of the treatment difference meets a predefined threshold for clinical relevance. This more complex case will not be discussed in detail in the present document.
* “High probability” of a treatment difference above target thresholds set based on clinical and/or commercial considerations. This approach is often used in early phase studies (e.g., proof of concept) to inform the decision to proceed with the development of the drug.

Given the multifaceted nature of the concept of success, the criteria outlined here necessarily involve a degree of simplification. For example, safety findings must also be considered when establishing the benefit-risk profile of a drug. Ultimately, the totality of evidence determines the success of a study. However, efficacy is generally the most critical component influencing the probability of success and is the one for which a quantitative assessment is more readily achievable. For this reason, these success criteria are the primary focus of this document. While more comprehensive definitions of study success are possible, they come at the cost of increased complexity in PoS evaluation.

The first step in evaluating the PoS of the study is to clarify the definition of success considered appropriate and relevant in the specific context. PoS evalution under both approaches is discussed in the following sections.

## Power and PoS

Most non-statisticians are familiar with statistical power in the context of clinical trials. It's related to PoS but differs in how it treats uncertainty, this is effectively the difference between conditional and unconditional probability.

Statistical power is the (conditional) probability of achieving statistical significance IF the true, underlying effect of the treatment is *exactly* equal to a specific value assumed during the design phase (the "target effect size").

**💡 Helpful Tip: Explaining Statistical Power** **(conditional probability):**

Explain, "You're all familiar with 'power' from our Phase 3 sample size calculations. That power figure (e.g., 80% or 90%) is *conditional* – it assumes the true treatment effect is *exactly* what we specified in the design (our target effect size). If that assumption is off, the actual chance of success can be different."

It’s important to advise non-statisticians that, while focusing on a single target effect size may be conceptually simpler, it has limitations. Power is highly sensitive to the accuracy of the assumed effect size and if the true underlying effect size is smaller, the actual chance of success will be lower than the calculated power suggests.

In contrast, PoS incorporates uncertainty into the estimate, which is a key point to communicate to non-statisticians. It’s essential they understand that PoS, often referred to as average power or assurance, accounts for a range of plausible true effect sizes and their associated likelihoods, based on prior knowledge or beliefs.

This prior knowledge about the plausible range and likelihood of the true effect size is formally captured in what is known as a prior distribution (FDA Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, 2010). This prior distribution is typically informed by data from earlier-phase studies (e.g., Phase 2), relevant published literature, results observed with similar drugs, or expert opinion. The PoS is then calculated by essentially averaging the conditional power across possible effect sizes represented in the prior distribution, with conditional power values weighted by the probability assigned to that specific effect size by the prior.

**💡 Helpful Tip: Explaining PoS (Unconditional probability):** To explain PoS simply, encourage non-statisticians to think of it as a kind of "average power," where the average is taken across each possible true effect size. Each effect size is weighted by a probability reflecting our belief in its likelihood, based on existing data or expert input. This results in a more realistic and robust estimate of the likelihood of success.

To illustrate the concept of PoS as average power, a simplified example may be built by considering only a few discrete values of the potential true effect size, instead of the entire range. Each effect size is assigned a probability based on prior knowledge, ensuring the total sums to 1. PoS is then calculated as a weighted average of power, using the prior probabilities as weights. By suitably choosing discrete values for the true effect size, the PoS calculated using this simplified method can often closely approximate the value obtained by integrating the power function over the prior distribution.

|  |  |  |
| --- | --- | --- |
| **True treatment effect** | **Prior probability** | **Power** |
| -1 | 5% | ~0% |
| 0 | 10% | 2.5% |
| 1 | 25% | 28.8% |
| 2 | 45% | 80.0% |
| 3 | 15% | 98.8% |
|  |  | **Average power = 58.3%** |

### Approaches to effectively communicate PoS

An Overlay Plot [left panel of plot below] is often the most intuitive and effective visual aids for communicating PoS. It plots two curves on the same graph:

* The *prior distribution*, representing beliefs about the probability of each true effect size.
* The *conditional power curve*, an S-shaped curve showing the probability of success of the planned trial, conditional on each possible effect size (i.e., the power).

This plot visually combines prior belief with expected trial performance. If the prior peaks in a region where power is high, the PoS will be high. Conversely, if the prior suggests the effect is likely small (where power is low), the PoS will be low. To enhance interpretability, a reference line can be added to represent the traditional power calculation (e.g., 80% power at an assumed effect size of 2 in the example). This helps bridge the gap between power and PoS.

Additionally a distribution plot [right panel of plot below] demonstrating the area under the curve for the prior distribution is also a useful visual tool when communicating the range of prior beliefs.

**💡 Helpful Tip: Leveraging Visuals like Overlay Plots** Advocate for using visuals. When showing an overlay plot, explain to your audience: "One curve shows what we *believe* about the drug's effect based on past data (our prior distribution). The other S-shaped curve shows how likely our *planned trial* is to succeed at each possible effect size (conditional power). The area where these curves align most closely gives us insight into PoS: it tells us how much of our belief lies in the range where the trial is likely to succeed". For the distribution plot (right panel), highlight that "the shaded area visually represents the PoS percentage – a simple takeaway for the overall chance of success." These plots are good for showing how PoS is derived by averaging conditional power across the prior.

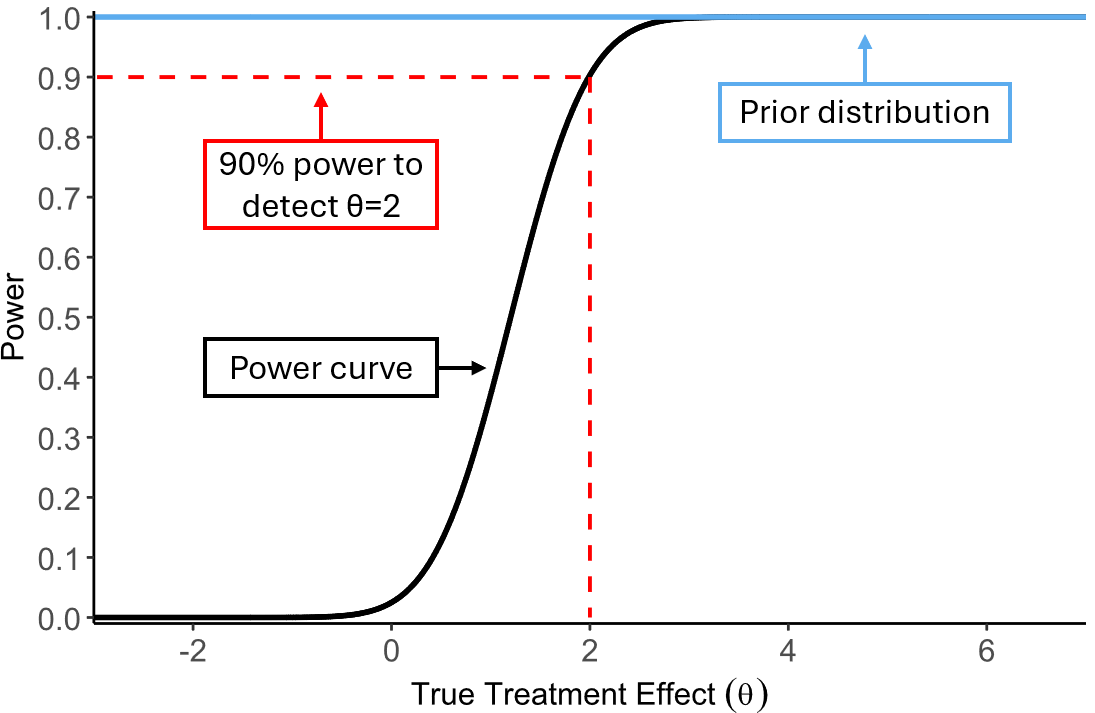
A diagram of a normal distribution

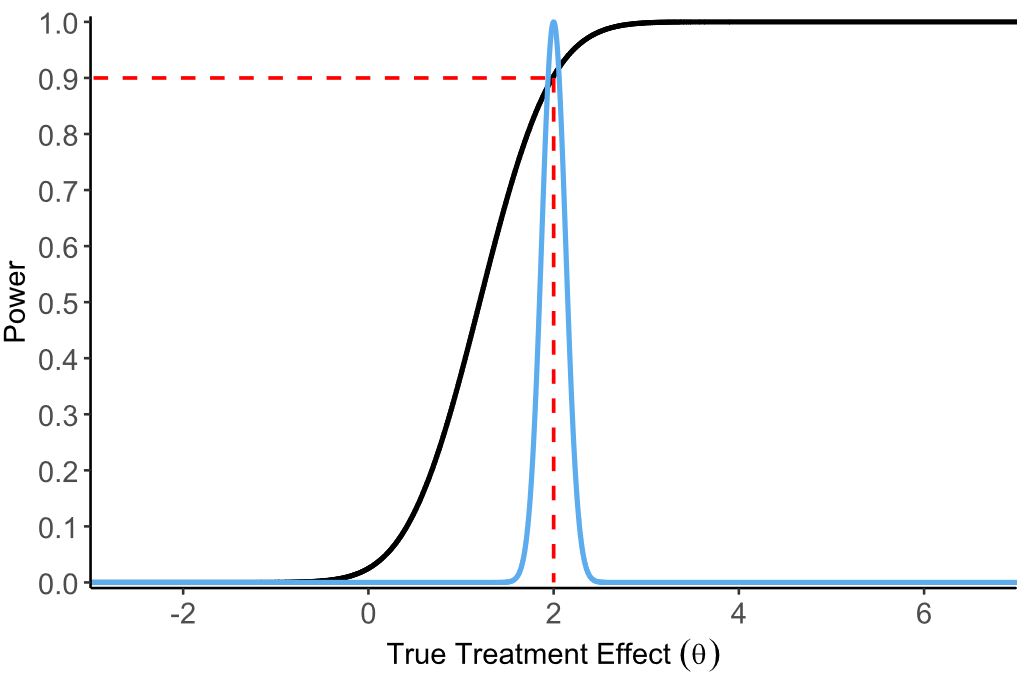
AI-generated content may be incorrect.

### **Properties of PoS**

There are several key properties of PoS that are important for non-statisticians to be aware of:

* **Impact of Prior Uncertainty:** The precision of prior distribution significantly influences the PoS (Morita et al, 2010).
  + **Uninformative Prior (e.g., very small Phase 2):** When existing data is limited and uncertain (wide confidence intervals of treatment effect), the prior distribution is spread out. The resulting PoS for the next trial might be close to 50%, reflecting that the existing evidence does not strongly support either success or failure, regardless of how large the next trial is.
  + **Informative Prior (e.g., large, convincing Phase 2):** If prior data is strong (narrow confidence intervals of treatment effect), the prior distribution is sharply focused. In this case, the PoS approaches the value of traditional power, reflecting greater confidence in the expected effect size.





**💡 Helpful Tip: Relating Prior Impact to Strength of Evidence Previous Phases**

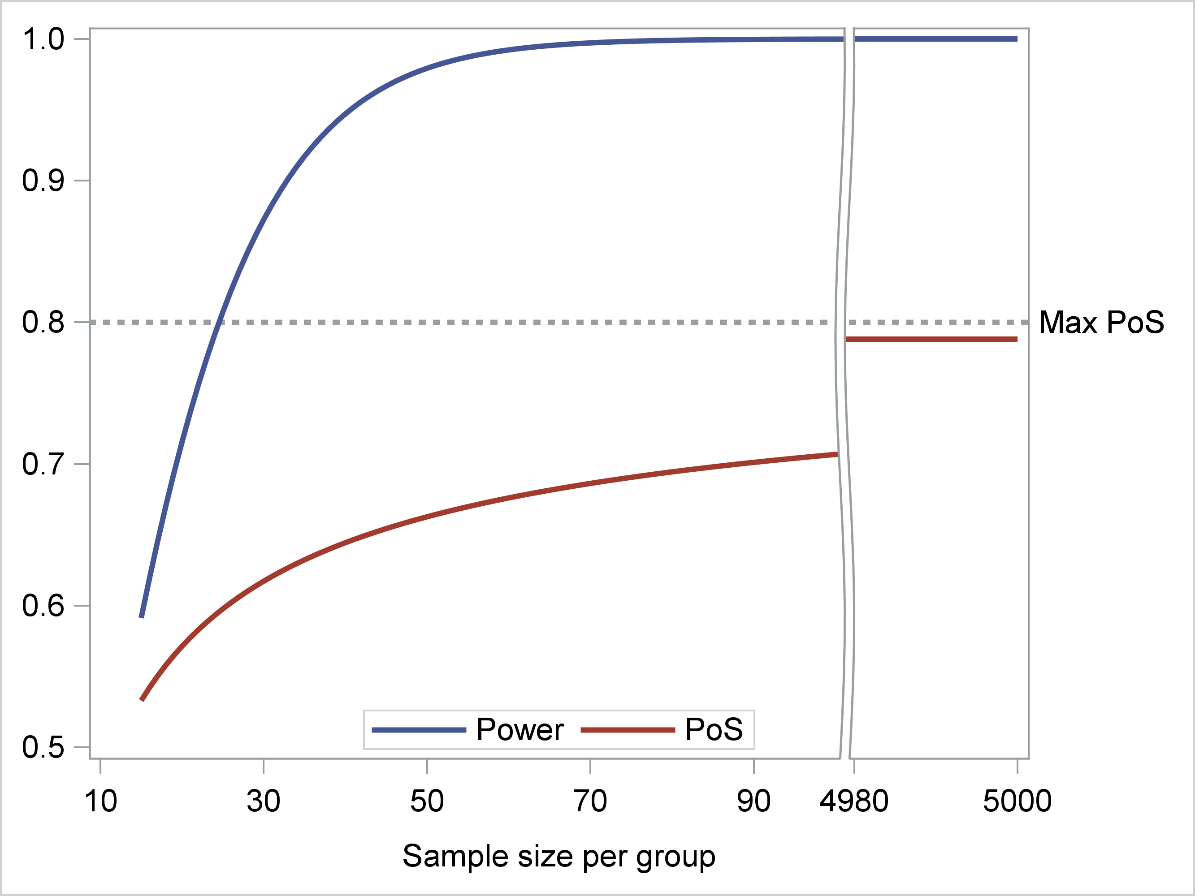
**Weak evidence**

"Because our prior knowledge is weak, we can't reliably predict the outcome of Phase 3. A PoS near 50% means that success and failure are equally likely.”

**Strong evidence**

"The Phase 2 results are clear and support the assumptions used in the power calculation. Therefore, the power is a reliable estimate of the PoS for Phase 3."

* **PoS Has a Ceiling (Boundedness):** This is a critical difference from traditional power. Power can approach 100% if the trial is made infinitely large (assuming the true effect is not zero). In contrast, PoS is bounded – it cannot exceed the prior probability that the treatment effect is favorable (e.g., greater than zero, or above a minimally important difference).
  + **Example to share:** Clarify that PoS can't ignore past doubts. Suppose prior results were encouraging but uncertain, with a wide confidence interval and suggesting a 20% chance the true effect is actually zero or negative (e.g., a one-sided p-value of 0.2). Even with a very large Phase 3 trial, you cannot eliminate the 20% chance that the treatment is ineffective. Therefore, the PoS for the next study is capped at 80% (1 - 0.20), regardless of sample size. See the example in the figure below.



**Sample size determination**

During the study design phase, when the Probability of Success (PoS) is considered in sample size determination, a joint graphical representation of power and PoS across varying sample sizes can be a valuable tool to support discussions with the study team. An example is shown in Figure X above. This visualization is particularly useful for demonstrating how sample size adjustments affect power and PoS differently, due to the ceiling effect inherent in PoS. In some scenarios, increasing the sample size may may substantially increase power but yield only a modest gain in PoS, raising questions about the justification for a larger study.

**💡 Helpful Tip: Impact of sample size on power and PoS** Always show how sample size effects both power and PoS. A significant increase in power with a larger sample size doesn’t automatically translate into a meaningful improvement in PoS. It’s essential to evaluate whether a bigger study truly adds value.

## Minimum Value (MV), Target Value (TV), and Go/No Go/Consider Regions

In clinical trials, particularly in Phase 2 trials, defining clear success criteria is crucial for making informed decisions about whether to continue developing a treatment. The Minimum Value (MV) is the lowest acceptable outcome for a trial to be considered successful, while the Target Value (TV) is the desired optimal outcome (NIH Example Target Product Profile). MV and the term lower reference value (LRV) can be also used interchangeably (Fewer et al, 2016).

These values are determined with input from various experts, including clinical, regulatory, and commercial teams, and are documented in the target product profile (TPP). Effectively communicating how likely the current trial is to meet these values is key to effective trial design.

It is assumed for this section that the non-statistical audience will be aware of MV and TV and that the choice of MV and TV are extremely important to the design. This section will focus on the communication of probabilities exceeding MV and TV during the trial design phase, along with communicating how these values can be used to establish Go/No Go/Consider regions to guide decision-making regarding the future development of an investigation treatment.

**MV and TV**

During the design stage of a trial it’s important to communicate the probabilities of exceeding the MV and TV set for the trial. Using graphs to show how the probability of exceeding MV and TV changes with different sample sizes can make these concepts more understandable. For instance, a line plot with sample size on the x-axis and probability on the y-axis can visually demonstrate the impact of study size.

Its extremely important when designing the study to set minimum probability thresholds for MV and TV that the study team are comfortable with. The minimum acceptable probability threshold may depend on therapeutic area, the unmet need, commercial opportunity and the companies appetite for risk. As a result the decision on the thresholds should involve a cross-functional discussion ensuring all perspectives are considered.

**💡 Helpful Tip: Discussing Likelihood of Hitting MV/TV** Guide the team to understand it's not just *if* the trial *can* hit the MV or TV, but *how likely* it is. Facilitate a discussion to set "comfort levels" (i.e. minimum probability thresholds) for achieving MV and TV (e.g., "We want at least an 80% chance of hitting our Minimum Value"). Explain that these desired probabilities influence study design, particularly sample size, and depend on various factors like therapeutic area and risk appetite.

**Go/No Go/Consider Regions**

Go/No Go/Consider Regions (Jiang et al, 2025) regions are used to categorize the trial outcomes based on the MV and TV:

* **Go Region:**
  + The Go Region can be easily communicated to non-statisticians as the range of scenarios where the trial results exceed the MV with the desired level of confidence.
* **No Go Region:**
  + The No Go Region can also be easily communicated to non-statisticians as the range of scenarios where the trial results fall below the TV with the desired level of confidence.
* **Consider Region:** The Consider Region can be communicated as the range of scenarios where the trial results neither (1) exceed the MV with the desired level of confidence to fall into the Go Region or (2) fall below the TV with the desired level of confidence to fall into the No Go Region.

Effective communication of these regions is very important, using graphical examples to illustrate the Go, No Go, and Consider regions can help non-statisticians visualize and understand these concepts better. Clear, simple explanations and visual aids can bridge the gap between statistical analysis and practical decision-making. An example of scenarios falling into Go, No Go and Consider regions can be seen in figure below (Fewer et al, 2016).

A diagram of a treatment effect

AI-generated content may be incorrect.

By assessing the probability of success (PoS) in the context of MV, TV, and Go/No Go/Consider regions, we can make more informed decisions about the future of a treatment. This approach helps identify potential risks and opportunities, ensuring that resources are used effectively and that development efforts are focused on the most promising treatments. It allows developers to assess the risk and potential of a drug candidate in the proposed trial. For instance, if the PoS to be in the Go region is very low then the team may want to reconsider the study design or whether to proceed with the study at all. Similarly if the trial has a high probability of falling in the Consider region then the development team need to be prepared for what additional information they will need to make a decision after falling in this region. Assessing the PoS of these regions allows for the assessment of multiple endpoints in the decision criteria.

**Communicating MV, TV, and Go/No Go/Consider Regions to Non-Statisticians**

* **Simulation**: Use simulation to demonstrate Go/No Go/Consider regions change depending on study design assumptions.
* **Visual Aids**: Use graphs and charts to illustrate the MV, TV, and Go/No Go/Consider regions. This can help non-statisticians visualize the concept and understand how the trial results relate to the decision criteria. The use of interactive graphs that show the impact of varying assumptions can help inform the team.
* **Examples**: Provide prior examples of how MV, TV, and Go/No Go/Consider regions have been used in actual clinical trials that are similar to your study. This helps to make the concepts more relatable and understandable.
* **Focus on the Implications**: Explain the implications of the trial results falling into each region. What does it mean for the drug's development if the results are in the Go region, the No Go region, or the Consider region?

**Graphical Approaches Communicating Go/No Go/Consider with Non-Statisticians:**

Go, No Go and Consider region plots (Lalonde plots) allow the study team to explore their study design over a range of treatment effects. They are particularly helpful communicating a range of treatment effects with non-statisticians as they:

* Provide Visual Clarity:
  + The plot uses clear colour-coded bands (green, yellow, red) to represent different decision outcomes. This makes it easy to understand at a glance whether a treatment is likely to be successful (Go), needs further consideration (Consider), or is unlikely to be effective (No Go).
  + Non-statisticians can quickly grasp the likelihood of success without needing to delve into detailed statistical calculations.
* Intuitive to Understand:
  + The Y-axis shows the probability, which is a familiar concept even to those without a statistical background. Seeing how these probabilities change with different treatment effects on the X-axis helps in understanding the overall effectiveness.
* Facilitates Decision Making:
  + By categorizing outcomes into Go, Consider, and No Go, the plot provides a straightforward way to interpret the results of extensive and complex simulations over a range of treatment effects.
* Easy to communicate:
  + These plots are excellent tools for communicating findings to stakeholders who may not have a deep understanding of statistics. They can see the potential outcomes and make informed decisions based on the visual representation.
  + By clearly showing the probabilities of different outcomes, the plot helps align expectations and facilitates discussions about the next steps in the decision-making process.

**Phase 2 Diabetes Example:**

Here we show an example diabetes study with the change in HbA1C as the primary endpoint. The study has 32 patients per arm in the active and placebo arms and a standard deviation of 1.2%. The probability has been calculated based on 10,000 simulations per treatment effect with the MV and TV equal to -0.5% and -1% respectively.

When communicating with non-statisticians it's important to highlight the following:

* This x-axis shows the difference in treatment effect between the new treatment and a placebo. Lower values means the treatment is more effective.
* The y-axis shows the probability of the treatment effect falling into one of the three categories.
* The coloured bands:
  + The green band “Go” represents the probability that the observed treatment effect is < -1%.
  + The yellow band “Consider” represents the probability that the observed treatment effect is <= -0.5% and > -1%.
  + The red band “No Go” represents the probability that the observed treatment effect > -0.5%.
* When the true treatment effect relative to placebo is a 1% improvement or greater then the probability of a Go decision is at least 50%.
* When the true treatment effect relative to placebo is 0.5% improvement or less then the probability of a No Go decision is at least 50%.
* When the true treatment effect relative to placebo is an improvement between 0.5% and 1% then the probability of a Consider decision is approximately 45%.

Lalonde plot (Fewer et al, 2016):

A graph showing different colored lines

AI-generated content may be incorrect.

**💡 Helpful Tip for Statisticians: Guiding Interpretation of Lalonde Plots**

When presenting a Lalonde plot, walk your audience through it:

* "The **bottom axis (X-axis)** shows different possible true effects of our drug – for instance, how much it might lower blood sugar compared to placebo. Remember, in this example, lower (more negative) values are better."
* "The **side axis (Y-axis)** shows the probability."
* "The **colored bands** are key. For any given true effect on the bottom, look up. The colors show the chances of our study results landing in a 'Go' (green), 'Consider' (yellow), or 'No-Go' (red) zone, based on our defined MV and TV and study design. This helps everyone see how likely each outcome is across a whole range of possibilities and discuss the implications for the study design."
* Point out specific interpretations from the plot, like "If the true effect is X, we have a Y% chance of a 'Go' decision".

## Using Probability of Success in Interim Analysis

Interim analyses are pre-planned assessments conducted during a clinical trial to evaluate the accumulating data. They allow for early decision-making regarding the continuation, modification, or termination of the trial based on emerging trends in safety and efficacy.

Measures conceptually similar to power and PoS (defined as average power) can play an important role in these interim analyses by providing a quantitative estimate of the likelihood of achieving the trial's objectives based on the available data.

* Conditional power can be seen as an analogous measure of power in an interim analysis setting. It is defined as the probability of rejecting the null hypothesis (e.g., of no test treatment effect) given the interim results and assuming a specific true effect size for the remainder of the trial.
* Predictive power or predictive probability of success (PPoS), if success is defined simply as the rejection of the null hypothesis, is an analogous measure of PoS in an interim analysis setting. It is defined as the conditional power averaged over a range of treatment effects, with weights based upon current belief about the effect size, represented by its posterior distribution. The posterior distribution of effect size can be based on interim data only (i.e., derived as non-informative prior + interim data) or on both pre-trial knowledge and interim results (informative prior + interim data).

**💡 Helpful Tip:** Explaining PoS-like Measures at Interim Looks:

* "**Conditional Power:** 'Given what we've seen so far, and assuming the drug's true effect is X for the rest of the trial, what's our chance of success?'."
* "**Predictive PoS (PPoS):** 'This is like updating our original PoS. Based on the early results, we revise our beliefs about the drug's effect (this is called a posterior distribution ) and recalculate an overall likelihood of trial success.' This helps the team decide whether to continue, stop, or modify the trial." Similar communication strategies as for pre-trial PoS apply here.

Due to these analogies, similar recommendations as above provided for PoS apply. If pre-trial beliefs about effect size are planned to be used in addition to interim data in PPoS calculation, it may be useful:

* To derive PPoS also using interim data only, to assess the sensitivity of PPoS to pre-trial knowledge.
* To perform a graphical comparison of the prior and posterior distributions of the effect size, to evaluate how interim data have modified expectations about treatment effect.

[EXAMPLE]

The above considerations apply to an unblinded setting, where the information about treatment assignment is available. Typically, in case of interim analyses, only the Data Monitoring Committee (DMC) members are unblinded. A different situation is that of a blinded observer only knowing the DMC recommendation to continue a study after an interim efficacy and/or futility analysis. If the pre-specified stopping rules are known, and assuming that the DMC adhered precisely to these rules, the PoS calculated before the trial can be updated taking into account DMC recommendation. [HIGHLIGHT THAT, DEPENDING ON THE STOPPING RULES, CONTINUING AFTER AN INTERIM FUTILITY ANALYSIS DOES NOT IMPLY A SUCCESS AT THE END, AND THE SAME FOR EFFICACY IN THE OPPOSITE WAY] [RELEVANCE OF GRAPHICAL REPRESENTATIONS TO BE EXPLORED (both power function and prior distribution of effect size needs to be updated] [EXAMPLE]

**Unblinded IA Example: Interim Analysis in a Diabetes Trial**

Consider a Phase 2 diabetes trial evaluating a new drug's effect on HbA1c reduction. An interim analysis is planned after half the patients have completed the study. The initial PoS for achieving a clinically meaningful reduction in HbA1c was estimated to be 60%.

After the interim analysis, the observed effect size is larger than expected, leading to an updated PoS of 80%. This information can be communicated to non-statisticians through distribution plots showing how the posterior distribution has shifted following the interim data.

A graph of a patient's distribution

AI-generated content may be incorrect.

**💡 Helpful Tip: Visualizing Changes in Belief Post-Interim** "This type of graph is excellent for showing how an interim look at the data can change our perspective. Explain that the 'Prior' distribution (often shown in one color, e.g., blue) was our belief about the drug's effect *before* seeing the mid-trial data. The 'Posterior' distribution (e.g., red) is our *updated* belief *after* incorporating the interim results. In an example like this, you'd point out how positive interim findings have shifted our belief towards a stronger drug effect, which in turn can boost the updated PoS (or PPoS)."

# References:

* P. Frewer, P. Mitchell, C. Watkins, J. Matcham, Decision-making in early clinical drug development, Pharm Stat. 15(3) (2016 May) 255-63.
* <https://doi.org/10.1002/pst.1746>.
* Lalonde RL, Kowalski KG, Hutmacher MM, Ewy W, Nichols DJ, Milligan PA, Corrigan BW, Lockwood PA, Marshall SA, Benincosa LJ, Tensfeldt TG, Parivar K, Amantea M, Glue P, Koide H. Miller R. Model-based Drug Development. Clinical Pharmacology & Therapeutics 2007; 82:21–32.
* Temple JR, Robertson JR. Conditional assurance: the answer to the questions that should be asked within drug development. Pharm Stat. 2021 Nov;20(6):1102-1111. doi: 10.1002/pst.2128.
* Epub 2021 May 7. PMID: 33960600; PMCID: PMC9291040.
* Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. Pharm Stat. 2018 Jul;17(4):301-316. doi: 10.1002/pst.1854.
* Epub 2018 Mar 30. PMID: 29603614.
* Saint-Hilary G, Robert V, Gasparini M. Decision-making in drug development using a composite definition of success. Pharm Stat. 2018 Sep;17(5):555-569. doi: 10.1002/pst.1870. Epub 2018 Jun 28. PMID: 29956453.
* U.S. Food and Drug Administration. The Use of Bayesian Statistics in Medical Device Clinical Trials: Guidance for Industry and Food and Drug Administration Staff , 2010. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm>
* Morita S, Thall PF, Müller P. Evaluating the Impact of Prior Assumptions in Bayesian Biostatistics. Stat Biosci. 2010 Jul 1;2(1):1-17. doi: 10.1007/s12561-010-9018-x. PMID: 20668651; PMCID: PMC2910452.
* Example Target Product Profile (TPP) for a Small Molecule Drug <https://seed.nih.gov/sites/default/files/2023-12/Example-TPP-Small-Molecule-Drug.pdf>
* Jiang, C. et al . Decision-making criteria and methods for initiating late-stage clinical trials in drug development from a multi-stakeholder perspective: a scoping review. Clin.
* Pharmacol. Ther. 117 , 978–988 (2025)

# Appendices

## Appendix 1: Interactive Tools

To support discussions and make PoS communication more effective, statisticians can leverage interactive tools such as the following:

- R Shiny apps (e.g., GOahead from BMS, <https://goaheadtool.shinyapps.io/GOaheadv10/>).

- Commercial solutions (e.g. PASS version XXXX, nQuery version XXXX).

## Appendix 2: Examples

## Early Phase PoS (Smart, 2025)

This example highlights how simulation and graphics can help convey the impact of priors, study design, and decision criteria on PoS.

The figure below helps convey probabilities based on a mixture prior, using an early-phase trial scenario. Here, the prior about the true treatment effect is bimodal—centered at 0 (indicating no effect) and 3 (indicating a clinically meaningful benefit). The decision criterion requires being 75% confident that the mean difference between UCB and placebo > 3.

Two panels compare the probability distribution of observed treatment effects for different sample sizes (N=30 vs. N=70 per arm):

* The green shaded region represents the area under the prior distribution where the study is expected to meet the decision criterion (i.e., PoS).
* The red region represents the area where the trial is expected to fail to meet the success criterion.

Key Communication Insights:

* With 30 patients per arm, the probability mass is slightly wider; there is considerable chance of both success and failure across the prior distribution.
* With 70 patients per arm, the narrower distribution leads to slightly more precise predications—success is more likely if the compound truly works, but also less likely if it doesn’t.
* This visualization clearly shows how the PoS is a function of both prior and sample size—and how increasing N reduces ambiguity but doesn’t always increase PoS

Such plots help decision-makers understand that increasing sample size does not always increase the PoS—especially when the prior includes a realistic chance of failure. It also underlines the importance of the prior, and how decision criteria (e.g., “75% sure > 3”) interact with study design and assumptions.

A red and green graph

AI-generated content may be incorrect.