COA-CCT Session III

Using a standardized estimand framework for medical product review and labeling: a case study

FDA Statistics Mallorie Fiero

FDA Clinical Chana Weinstock

EORTC SISAQOL Madeline Pe

Panelists

- Andrea Ferris – Patient advocate
- Sigrid Klaar – European regulatory and payer perspective
- Alicyn Campbell – Industry
- Surya Singh – Domestic payer
- David Cella – Academic psychometrician
- Kim Cocks – Academic statistician

Take Home Messages

- There is a need for more **well-defined research objectives** that can be matched with appropriate statistical methods
  - **Estimand framework** is an organized approach to construct a well-defined endpoint

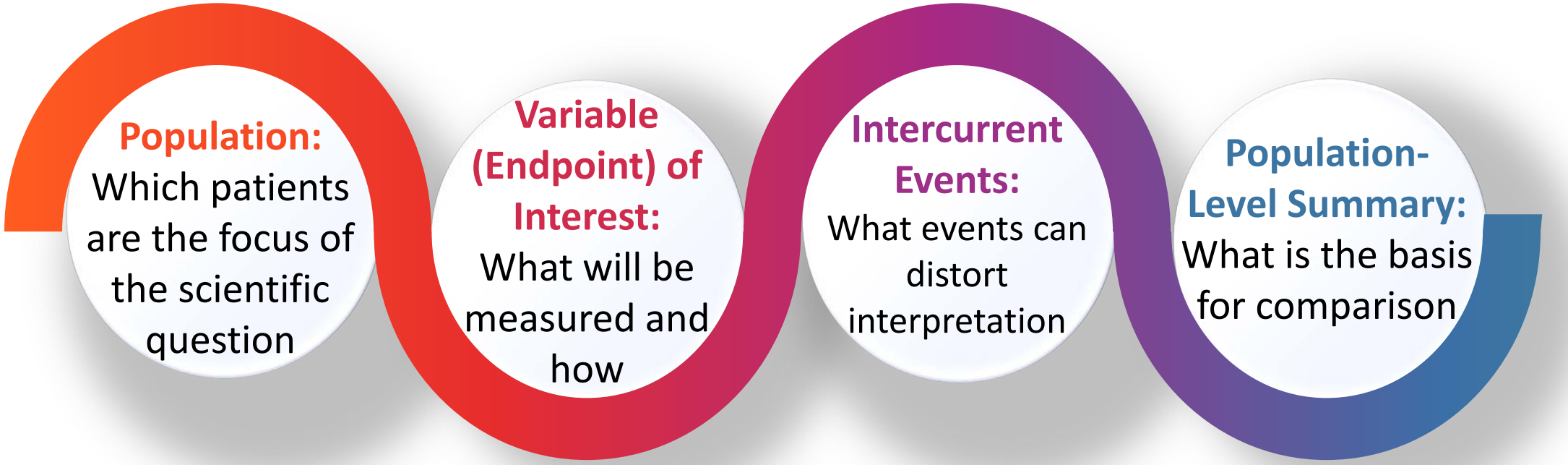
- Lack of superiority (e.g., $p > 0.05$) **does not mean** equivalence

- There is **no one best way** to evaluate patient experience, but **standard principles and analyses** must be developed

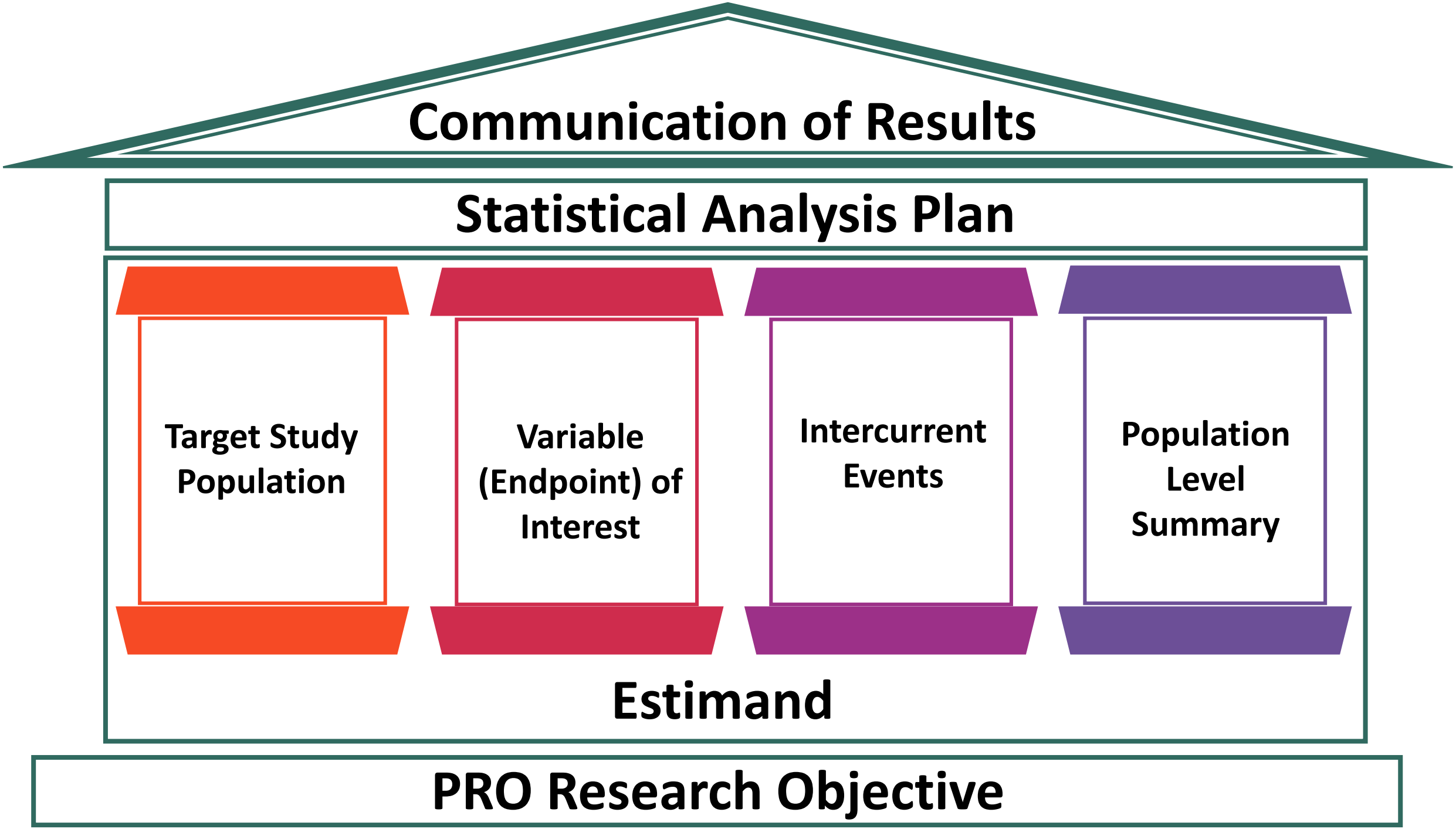
Session Outline

- Highlights of estimand framework
- **Research Objective 1: Supporting a marketing claim**
  - Panel discussion
  - Audience Q&A
  - Summary
- Mini-break (15 minutes)
- **Research Objective 2: Describing patient perspective on treatment**
  - Panel discussion
  - Audience Q&A
  - Summary
- Concluding remarks

Estimand Framework: Organized Approach to Construct a Well-Defined Endpoint



Estimand: Target of estimation to address a trial's scientific question of interest


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

DISCLAIMER

These case studies are **not an endorsement** of a singular study design, outcome, analysis, or visualization; rather it's meant to demonstrate how FDA may perceive physical function data in oncology

Two Broad Research Objectives

- **Research Objective 1: Supporting a marketing claim**
  - Conclusions regarding comparisons between treatment arms
  - *A-priori* hypothesis is needed
  - Statistical testing – correction for multiple testing is needed

- **Research Objective 2: Describing patient perspective on treatment**
  - No comparisons between treatment arms (e.g., CTCAE)
  - No *a-priori* hypothesis is needed
  - Descriptive/exploratory – multiple testing may be less of an issue

Case Study Clinical Scenario

- **Scenario**
  - Metastatic ER/PR+ HER2- breast cancer after progression on 1st line therapy

- **Epidemiology and Disease Information**
  - Breast cancer has heterogeneous disease symptoms and many women will be asymptomatic at baseline, even in the 2nd line setting
  - 2nd line prior studies have shown a median OS of 2-2.5 years with 2nd line hormone therapy alone and a median PFS of approximately 10-12 months

- **Treatment Goal**
  - Addition of targeted therapy to hormonal agent will improve PFS by 6-8 months
  - Combination is expected to add symptomatic toxicity

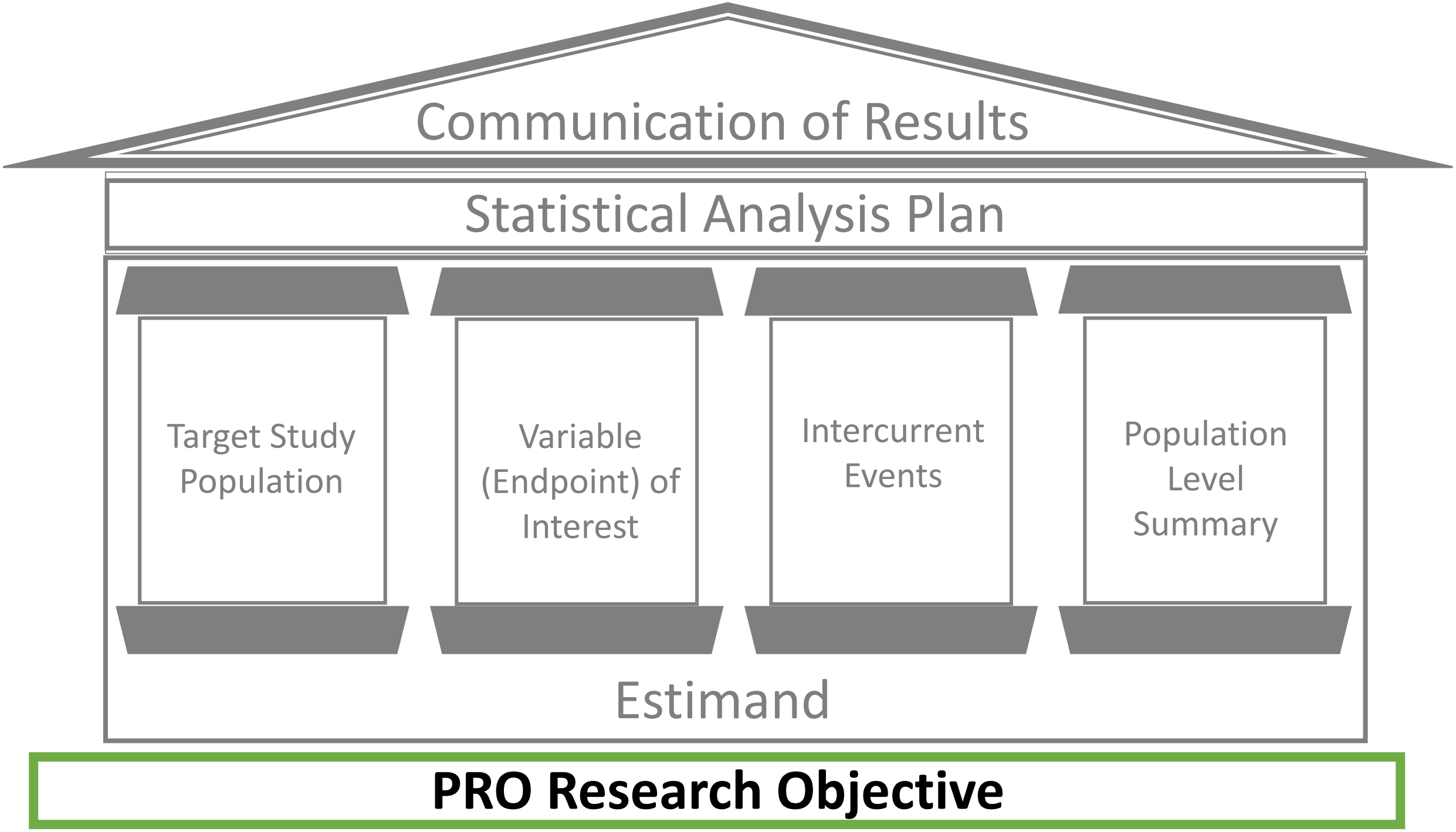
Source: https://www.uptodate.com/contents/treatment-approach-to-metastatic-hormone-receptor-positive-her2-negative-breast-cancer-endocrine-therapy-and-targeted-agents

Case Study Clinical Scenario

- **Study Design:** Randomized controlled trial
  - Treatment: SoC + oral targeted investigational agent
  - Control: SoC + placebo

- **Expected Outcomes**
  - Expected Efficacy: 6-8 month PFS benefit
    - OS may be impacted due to crossover
  - Expected Safety: Symptomatic toxicities including diarrhea, fatigue and rash greater on investigational arm

- **Population Assumptions**
  - Population is generally high functioning (ECOG 0 or 1)
  - Percentage of the population is symptomatic (from disease) at baseline


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define PRO Scientific Research Question
A Priori

PRO Research Objective

Superior benefit in physical function (PF) for the investigational arm compared to the control arm in the ITT population at Week 28

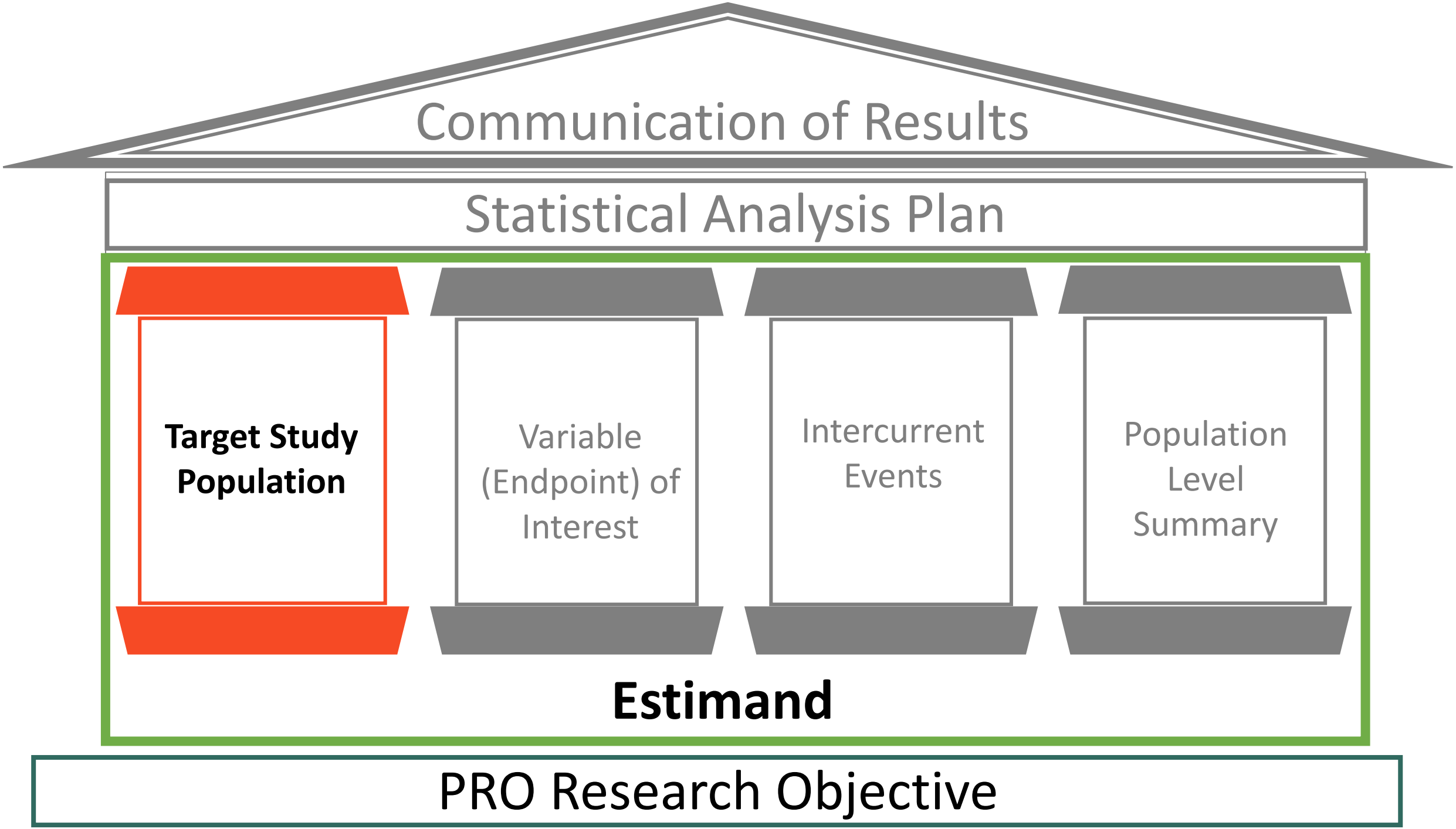
Scientific Research Question

What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Superiority vs. Non-inferiority/Equivalence Should be Pre-Specified

- **Inappropriate to conclude "no worsening" when there is a non-significant test of superiority (e.g., p > 0.05)**
  - Small sample size → wide confidence intervals → not likely to demonstrate superiority
  - PRO not sensitive to change

- Non-inferiority/equivalence challenges
  - Pre-specify meaningful non-inferiority/equivalence margin
  - Sample size often much larger than superiority trial
  - Poor study quality → bias towards equality
    - Missing data
    - Lack of compliance with treatment


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Target Study Population Based on Research Question *A Priori*

Scientific Research Question

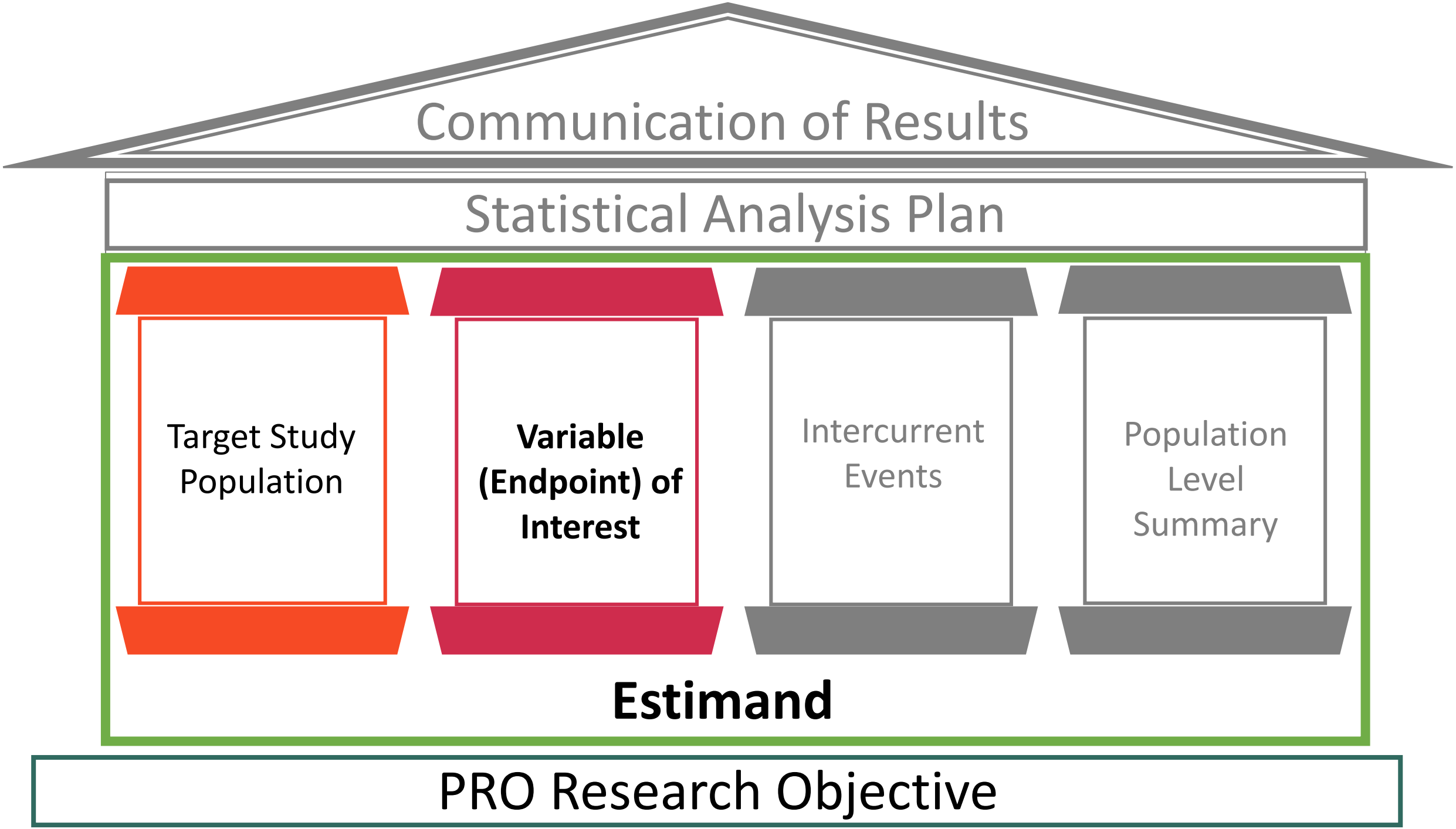
What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Target Study Population

Intent-to-treat (ITT) population

Defining the Target Study Population: Considerations

Target study population (examples)	
• ITT • Safety: All patients who received at least one dose of drug, regardless of randomization	• Analysis populations are often defined based on their availability of PRO data ➢ All patients who are eligible for PF PRO assessment ➢ Completed baseline PF assessment ➢ Completed baseline and at least one post-baseline assessment ➢ Any PF PRO data


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Variable (Endpoint) of Interest Based on Research Question *A Priori*

Scientific Research Question

What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Variable of Interest

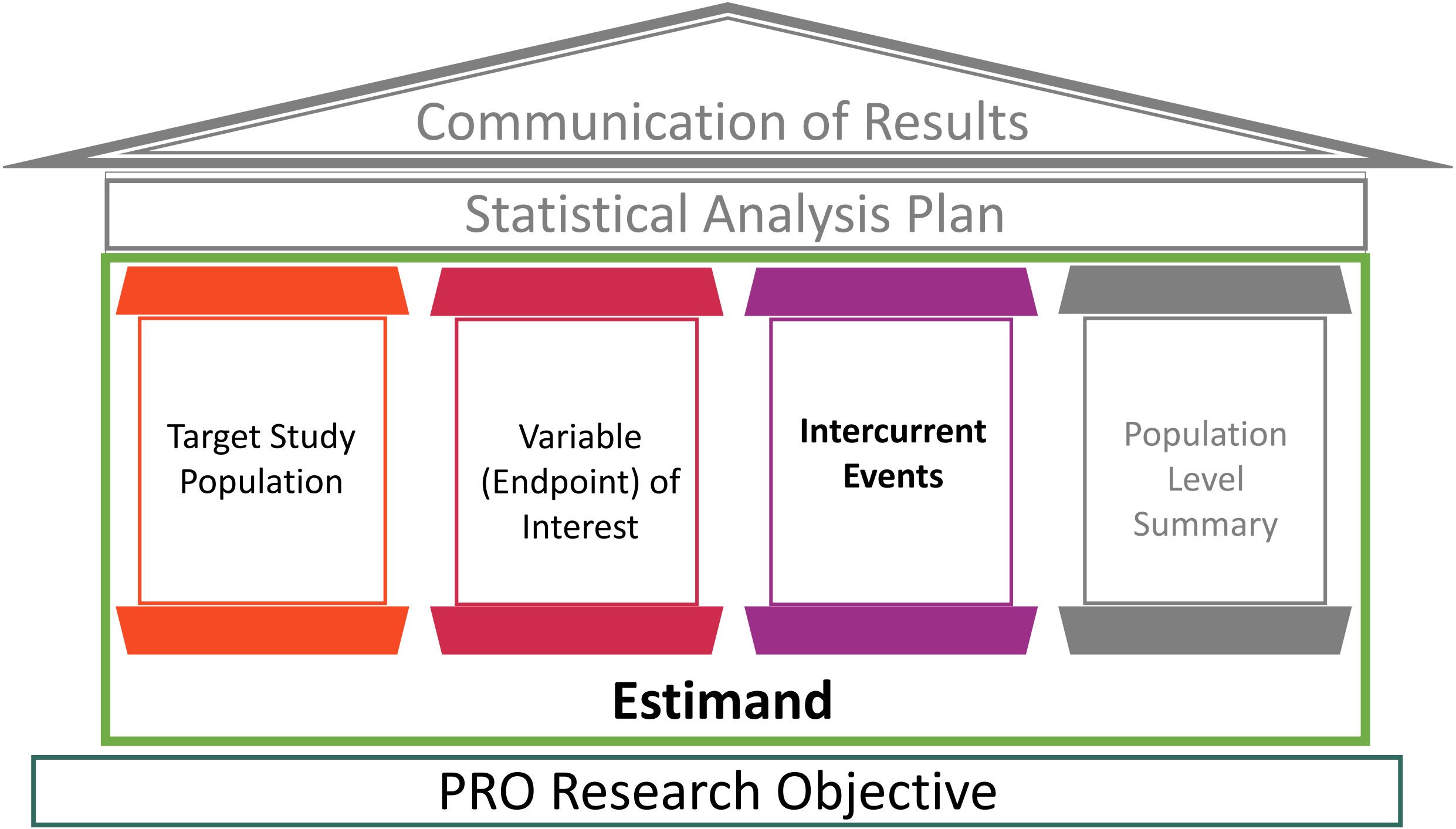
Change from baseline in PF score using well-defined measurement tool at Week 28

Defining the Variable (Endpoint) of Interest: Considerations

Concepts (examples)	Measurement tool qualities
• Physical function • Pain	• Well-defined • Reliable • Validated • Sensitive

Defining the Variable (Endpoint) of Interest: Considerations

Endpoint type	Analysis time point
• Time to event • Proportion with event at time *t* • Intensity/magnitude of event(s) at time *t* • Overall PRO score over time • Response patterns/profiles (longitudinal)	• Specific time point • Over time (specify time frame)


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

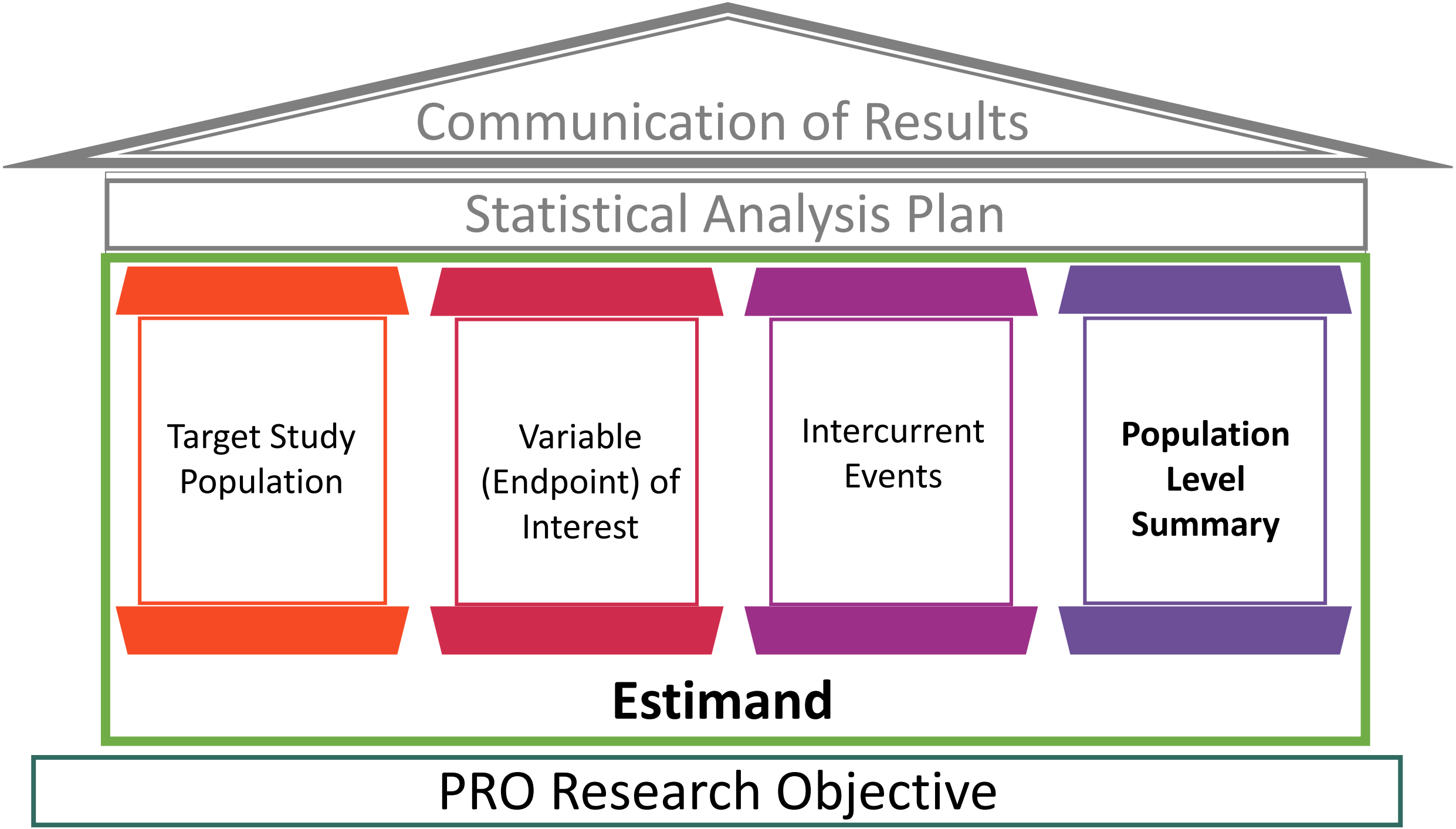
Address Intercurrent Events in Alignment with Research Question

Scientific Research Question
What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Intercurrent event	Handling of intercurrent event
• Death	PF not collected after intercurrent event occurs
• Discontinuation of treatment • Disease progression	PF collected regardless of whether intercurrent event occurs

Addressing Intercurrent Events: Considerations

Intercurrent events (examples)	Handling intercurrent events
• Death • Progression • Discontinuation due to adverse event • Taking subsequent therapy beyond discontinuation • Use of rescue medication or therapy • Hospitalization • Transplantation • Non-adherence	• There are multiple ways to handle intercurrent events • Pre-specify handling of intercurrent events in alignment with research question


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Population Level Summary Based on Research Question *A Priori*

Scientific Research Question

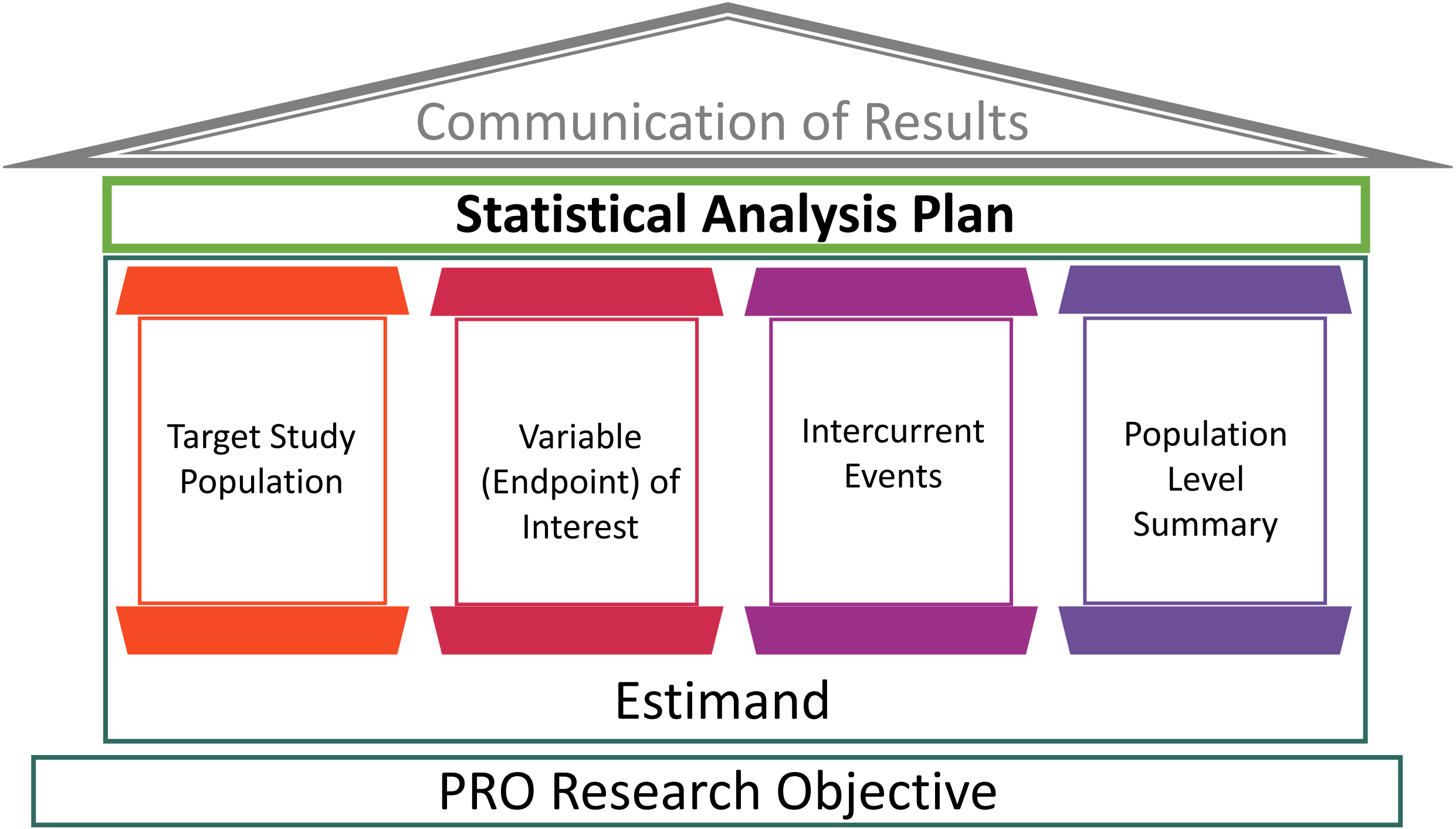
What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Population Level Summary

Least squares (LS) mean change from baseline in PF score at Week 28: Difference from control arm (95% confidence interval)

Defining the Population Level Summary: Considerations

Population level summary (examples)	Clinical relevance
• Median time to event, hazard ratio • Proportion of patients with event at time *t* • Mean change at time *t* • Mean overall PRO score over time (e.g., mean area under the curve) • Mean longitudinal profile	**Clinically relevant thresholds** • Within-individual change **Estimate** • Within-group mean change • Between-group difference

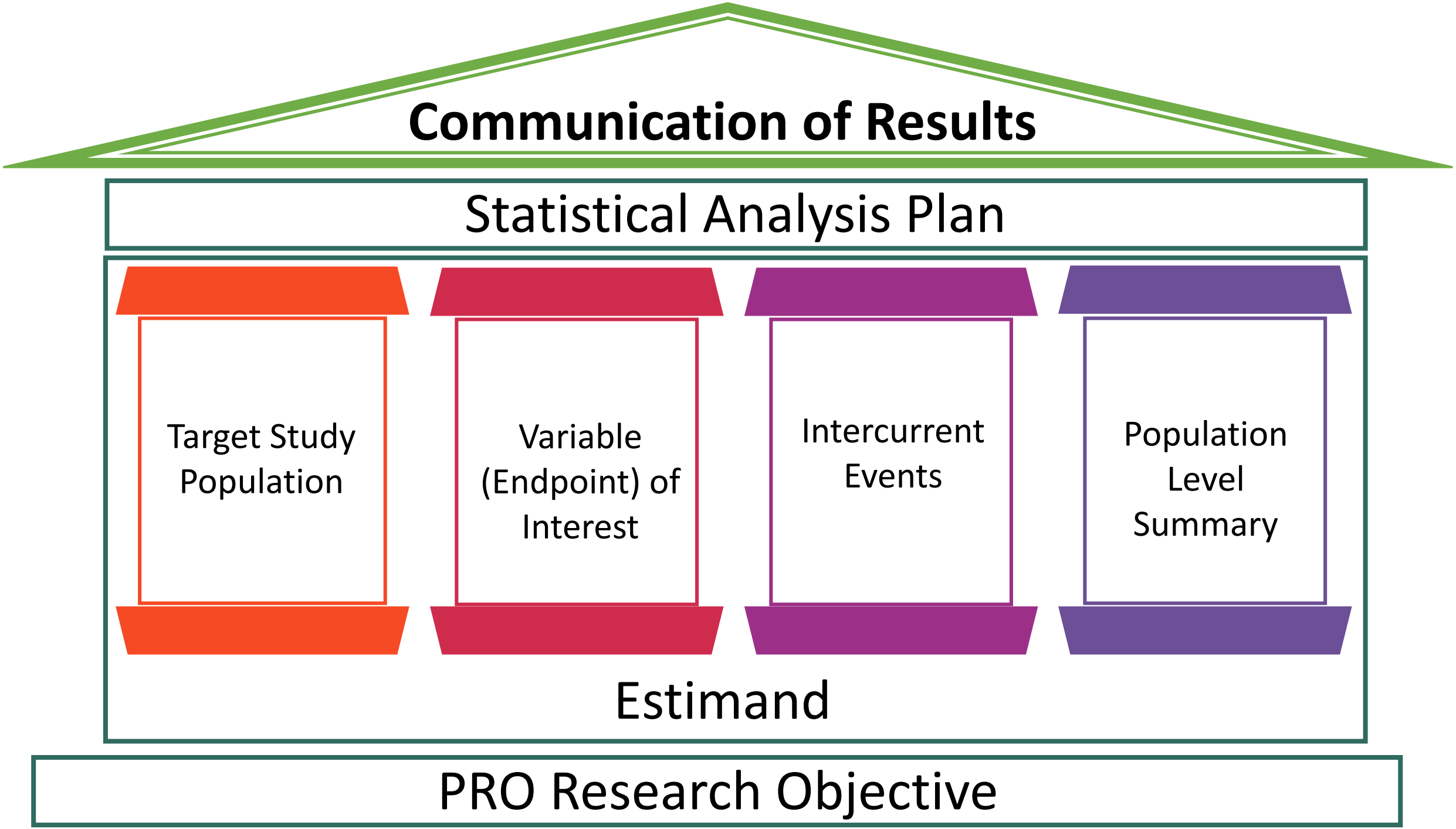

Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Scientific Research Question

What is the mean change from baseline in PF score at Week 28 among patients in the investigational arm compared to the control arm?

Statistical Analysis Plan

- **Efficacy endpoints**
  - Primary endpoint: PFS
  - Secondary endpoint: Mean change from baseline in PF score at Week 28
- **Analysis of mean change from baseline in PF**
  - Mixed models for repeated measurements (MMRM) in the ITT population
    - (Appropriate missing data assumption?)
  - Handling intercurrent events:
    - PF assessments will continue until date of death
    - PF data will be included regardless of progression or treatment discontinuation
- **Multiplicity**
  - Hierarchical testing plan


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Communication of Results

	Parameter	Treatment N = 198	Control N = 201
PF at Baseline	N	197	199
	Mean (SD)	70.4 (19.9)	74.0 (18.4)
PF at Week 28	N	178	181
	Mean (SD)	75.1 (16.2)	62.7 (15.7)
Change From Baseline in PF at Week 28	**LS Mean (95% CI)**	**4.6 (0.1, 9.1)**	**-10.6 (-15.7, -6.0)**
	Difference from control (95% CI)	**15.2 (8.7, 21.7)**	
	P-value	**< 0.0001**	

- Fabricated data
- Descriptive statistics and visualizations should also be performed for interpretation of within-individual change

Summary of Where Discussion Started
Research Objective 1: Supporting a Marketing Claim

Estimand attributes	Decisions to better define research objectives
Target population	ITT
Variable of interest	Change from baseline in PF score at Week 28
Handling of intercurrent event	
• Death	PF not collected after intercurrent event occurs
• Disease progression • Treatment discontinuation	PF collected regardless of whether intercurrent event occurs
Population level summary	LS mean change from baseline in PF score at Week 28: Difference from control arm (95% CI)

These case studies are **not an endorsement** of a singular study design, outcome, analysis, or visualization; rather it's meant to demonstrate how FDA may perceive physical function data in oncology

Panel Discussion

What are some considerations in assessing whether change in physical functioning is **clinically meaningful for patients** in the treatment arm?

	Parameter	Treatment N = 198	Control N = 201
Physical Function at Baseline	N	197	199
	Mean (SD)	70.4 (19.9)	74.0 (18.4)
Physical Function at Week 28	N	178	181
	Mean (SD)	75.1 (16.2)	62.7 (15.7)
Change From Baseline in Physical Function at Week 28	**LS Mean (95% CI)**	**4.6 (0.1, 9.1)**	**-10.6 (-15.7, -6.0)**
	Difference from control (95% CI)	**15.2 (8.7, 21.7)**	
	P-value	**< 0.0001**	

* Fabricated data

Questions From the Audience

- **Co-Moderators**
  - Mallorie Fiero – FDA statistician
  - Chana Weinstock – FDA clinician
  - Madeline Pe – SISAQOL

- **Panelists**
  - Andrea Ferris – Patient advocate
  - Sigrid Klaar – European regulatory and payer perspective
  - Alicyn Campbell – Industry
  - Surya Singh – Domestic payer
  - David Cella – Academic psychometrician
  - Kim Cocks – Academic statistician

Additional Panel Discussion Questions

1. Can you comment on how we handled intercurrent events? Should we assess for PF regardless of progression or discontinuation?

Intercurrent event	Handling of intercurrent event
• Death	PF not collected after intercurrent event occurs
• Discontinuation of treatment • Disease progression	PF collected regardless of whether intercurrent event occurs

2. Do you have additional considerations for the framework of including a PRO endpoint to **support a comparative claim**?

Take Home Messages

- There is a need for more **well-defined research objectives** that can be matched with appropriate statistical methods
  - **Estimand framework** is an organized approach to construct a well-defined endpoint

- Lack of superiority (e.g., $p > 0.05$) **does not mean** equivalence

- There is **no one best way** to evaluate patient experience, but **standard principles and analyses** must be developed

BREAK

DISCLAIMER

These case studies are **not an endorsement** of a singular study design, outcome, analysis, or visualization; rather it's meant to demonstrate how FDA may perceive physical function data in oncology

Two Broad Research Objectives

- **Research Objective 1: Supporting a marketing claim**
  - Conclusions regarding comparisons between treatment arms
  - *A-priori* hypothesis is needed
  - Statistical testing – correction for multiple testing is needed

- **Research Objective 2: Describing patient perspective on treatment**
  - No comparisons between treatment arms (e.g., CTCAE)
  - No *a-priori* hypothesis is needed
  - Descriptive/exploratory – multiple testing may be less of an issue

Case Study Clinical Scenario

- **Scenario**
  - Metastatic ER/PR+ HER2- breast cancer after progression on 1st line therapy

- **Epidemiology and Disease Information**
  - Breast cancer has heterogeneous disease symptoms and many women will be asymptomatic at baseline, even in the 2nd line setting
  - 2nd line prior studies have shown a median OS of 2-2.5 years with 2nd line hormone therapy alone and a median PFS of approximately 10-12 months

- **Treatment Goal**
  - Addition of targeted therapy to hormonal agent will improve PFS by 6-8 months
  - Combination is expected to add symptomatic toxicity

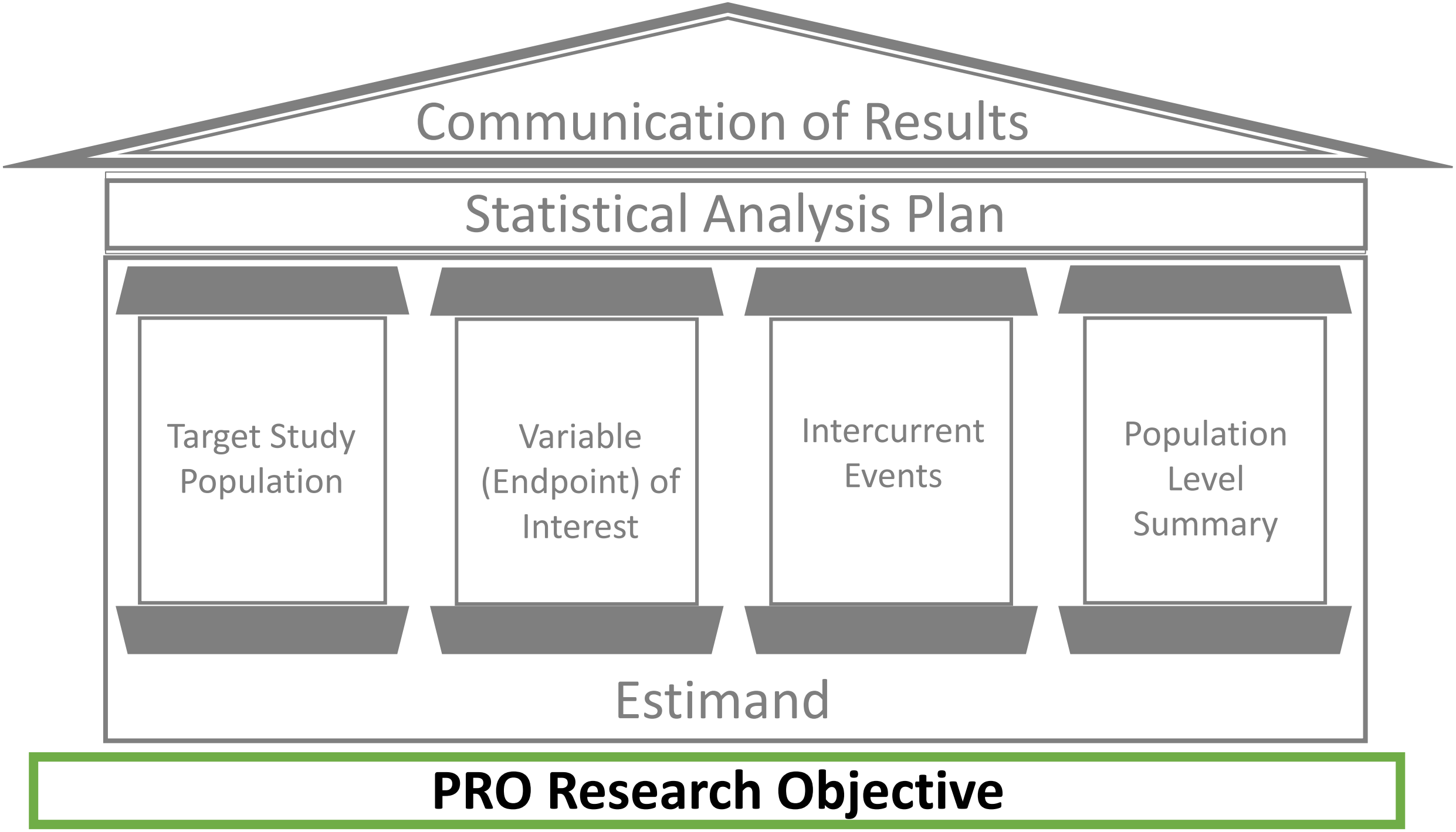
Source: https://www.uptodate.com/contents/treatment-approach-to-metastatic-hormone-receptor-positive-her2-negative-breast-cancer-endocrine-therapy-and-targeted-agents

Case Study Clinical Scenario

- **Study Design:** Randomized controlled trial
  - Treatment: SoC + oral targeted investigational agent
  - Control: SoC + placebo

- **Expected Outcomes**
  - Expected Efficacy: 6-8 month PFS benefit
    - OS may be impacted due to crossover
  - Expected Safety: Symptomatic toxicities including diarrhea, fatigue and rash greater on investigational arm

- **Population Assumptions**
  - Population is generally high functioning (ECOG 0 or 1)
  - Percentage of the population is symptomatic (from disease) at baseline


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

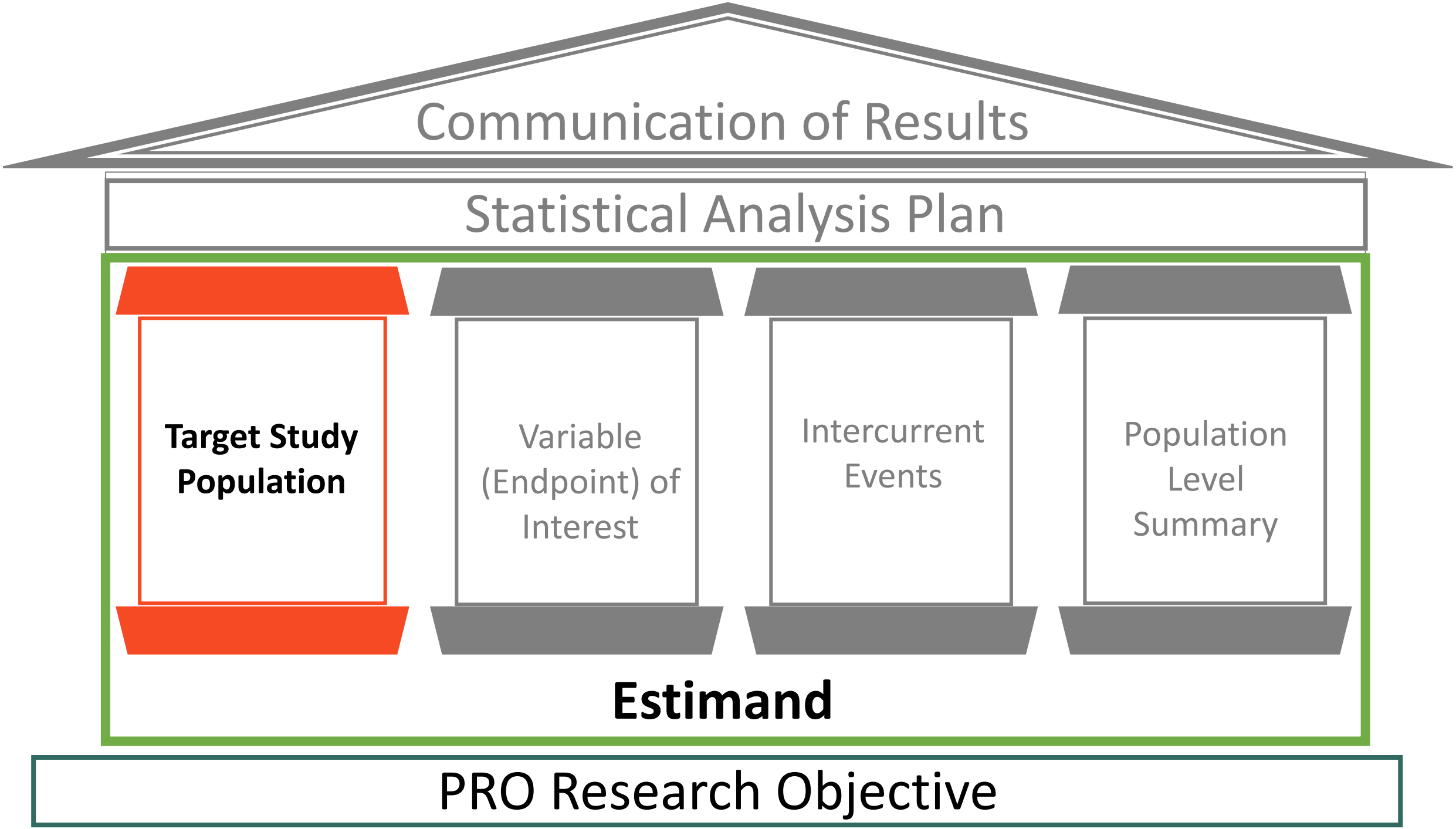
Define PRO Scientific Research Question *A Priori*

PRO Research Objective

Characterize physical function on investigational treatment

Scientific Research Question

Among patients on treatment, what proportion at least maintained their physical functioning?


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Target Study Population Based on Research Question *A Priori*

Scientific Research Question

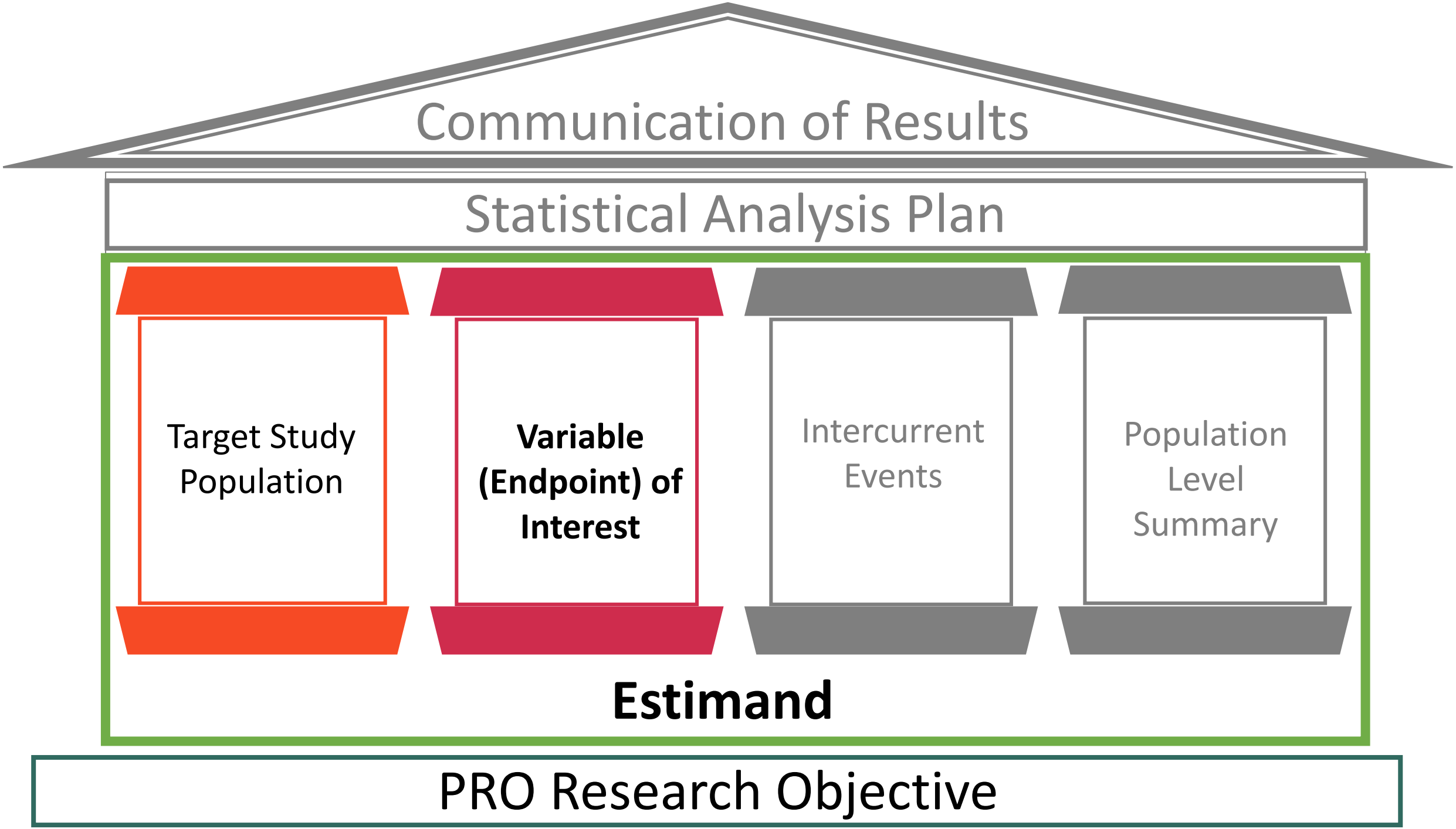
Among patients on treatment, what proportion at least maintained their physical functioning?

Target Study Population

Patients who received at least one dose of the drug + completed baseline PF assessment + on treatment

Defining a Target Study Population: Considerations

Target study population (examples)	
• ITT • Safety: All patients who received at least one dose of drug, regardless of randomization	• Populations are often defined based on their availability of PRO data ➢ All patients who are eligible for PF PRO assessment ➢ Completed baseline PF assessment ➢ Completed baseline and at least one post-baseline assessment ➢ Any PF data


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Variable (Endpoint) of Interest Based on Research Question *A Priori*

Scientific Research Question

Among patients on treatment, what proportion at least maintained their physical functioning?

Variable of Interest

At every assessment point until end of treatment, patients meeting pre-specified criteria* for PF maintenance/improvement using a fit-for-purpose measurement tool

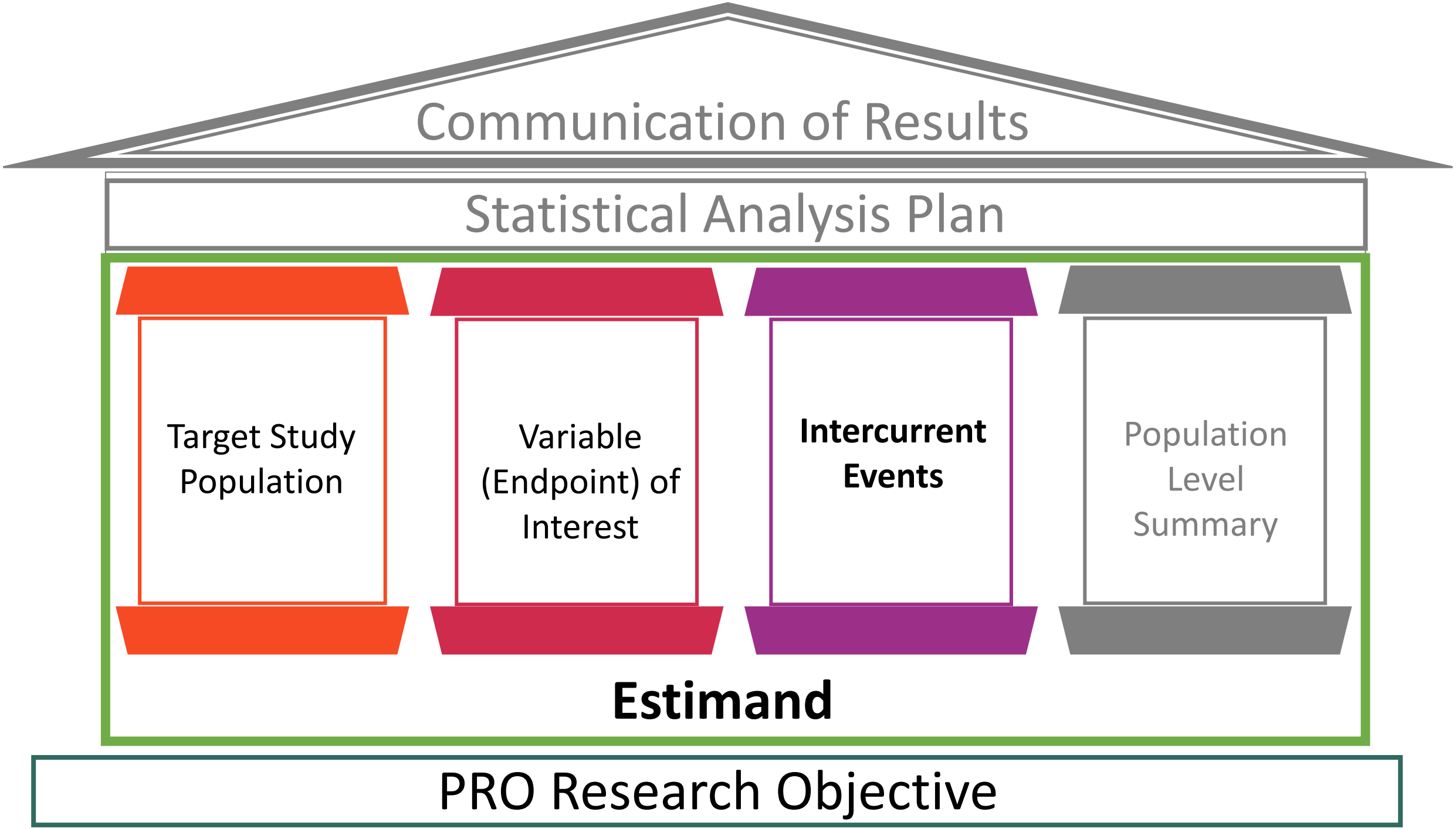
*Clinically relevant within-patient threshold for maintenance and improvement should be pre-defined

Defining a Variable (Endpoint) of Interest: Considerations

Concepts (examples)	Measurement tool qualities	Within treatment arm assumption
• Physical function • Pain	• Well-defined • Reliable • Validated • Sensitive	• Worsening • Maintenance • Improvement • No directionality assumption

Defining a Variable (Endpoint) of Interest: Considerations

Endpoint type	Analysis time point
• Time to event	• Specific time point
• Proportion with event at time *t*	• Over time (specify time frame)
• Intensity/magnitude of event(s) at time *t*	
• Overall PRO score over time	
• Response patterns/profiles (longitudinal)	


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

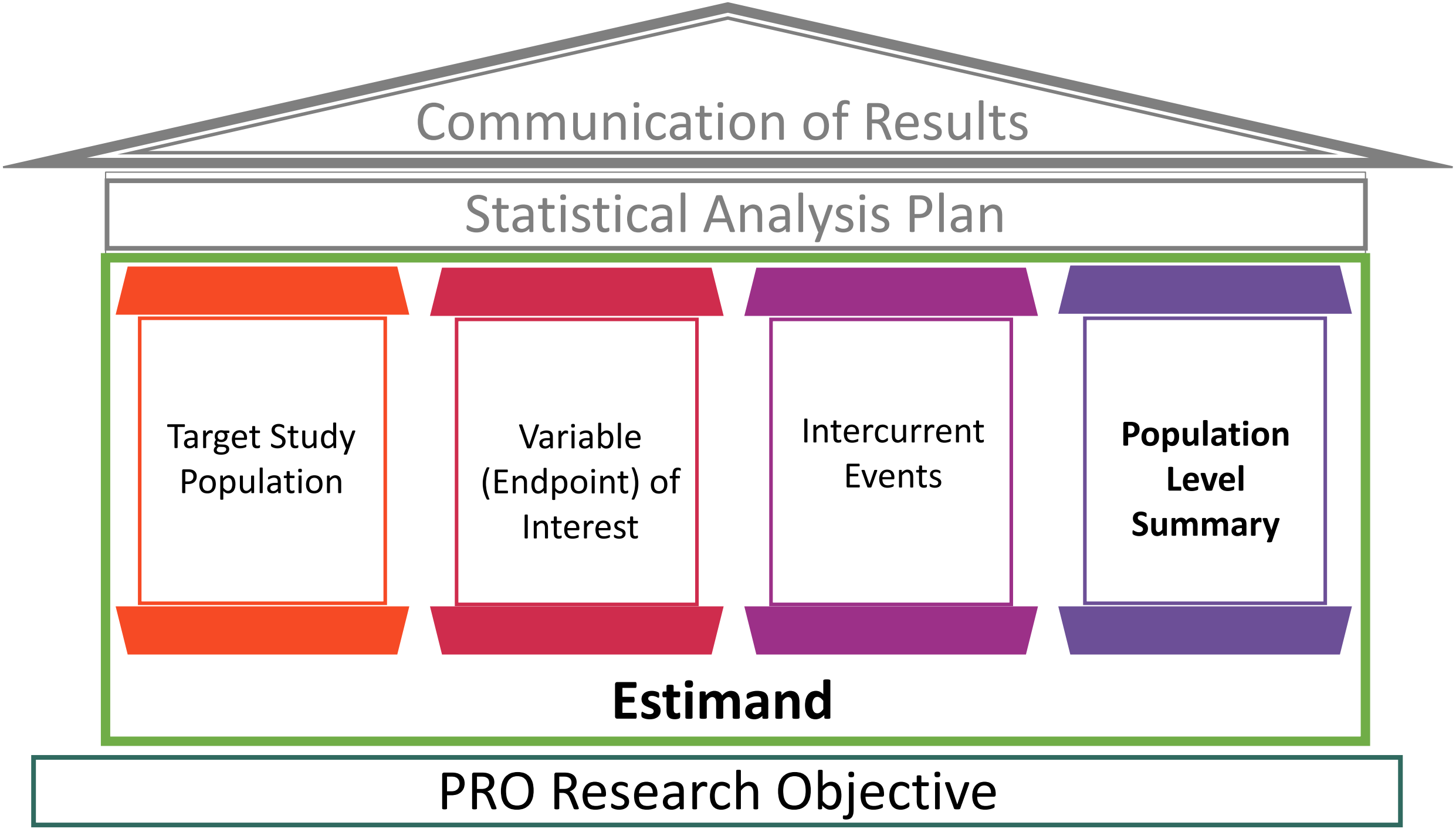
Address Intercurrent Events in Alignment with Research Question

Scientific Research Question
Among patients on treatment, what proportion at least maintained their physical functioning?

Intercurrent event	Handling of intercurrent event
• Death • Discontinuation of treatment • Disease progression	Patient dropped from analysis population after intercurrent event occurs

Addressing Intercurrent Events: Considerations

Intercurrent events (examples)	Handling intercurrent events
• Death • Progression • Discontinuation due to adverse event • Taking subsequent therapy beyond discontinuation • Use of rescue medication or therapy • Hospitalization • Transplantation • Non-adherence	• There are multiple ways to handle intercurrent events • Pre-specify handling of intercurrent events in alignment with research question


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Define Population Level Summary Based on Research Question *A Priori*

Scientific Research Question

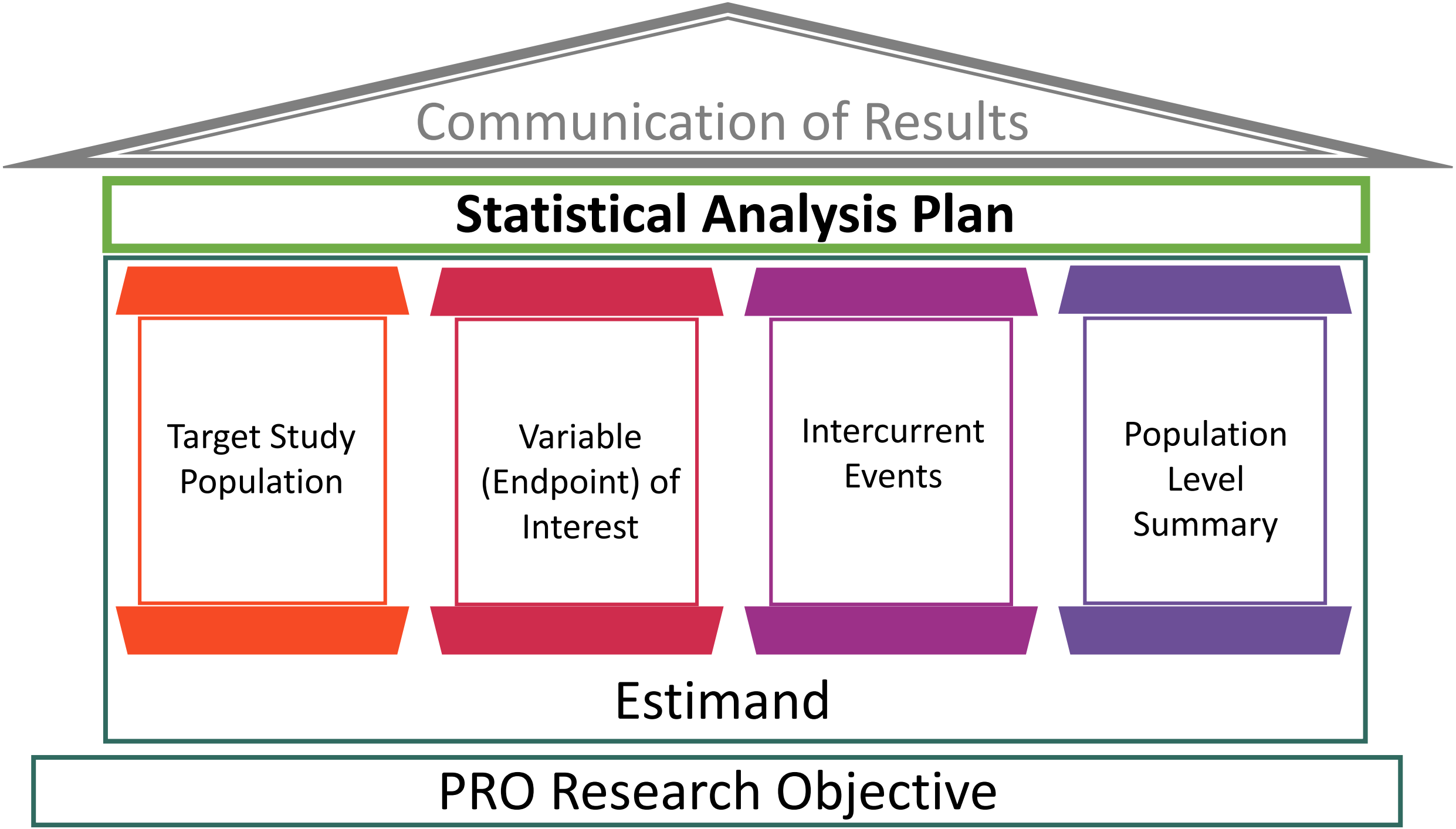
Among patients on treatment, what proportion at least maintained their physical functioning?

Population Level Summary

Proportion of on-treatment patients who maintained/improved PF

Defining a Population Level Summary: Considerations

Population level summary (examples)	Clinical relevance
• Median time to event, hazard ratio • Proportion of patients with event at time *t* • Mean change at time *t* • Mean overall PRO score over time (e.g., mean area under the curve) • Mean longitudinal profile	• Within-individual change • Within-group mean change • Between-group difference

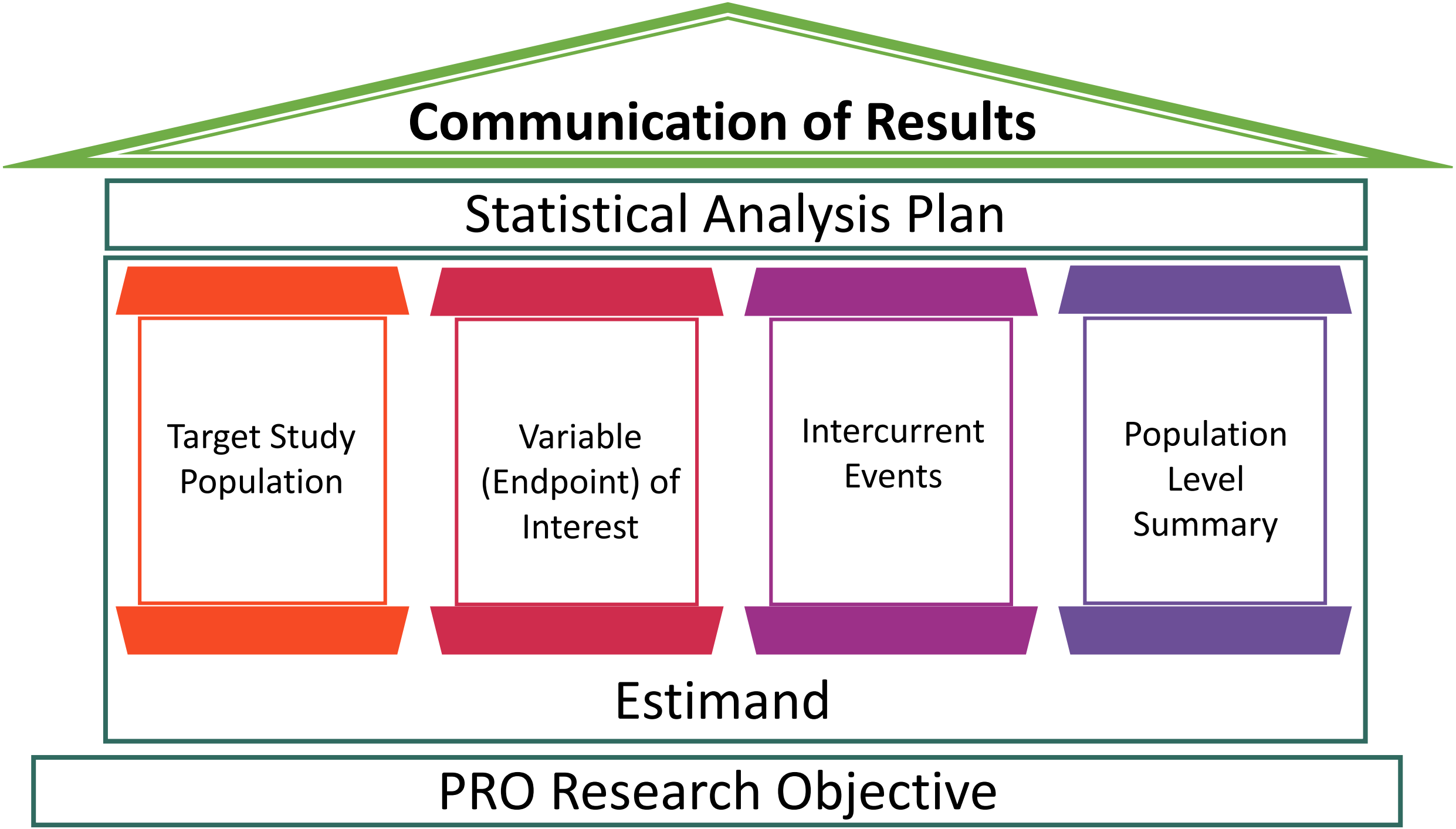

Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Scientific Research Question

Among patients on treatment, what proportion at least maintained their physical functioning?

Statistical Analysis Plan

- Proportion of patients who maintained or improved PF while on treatment will be **summarized descriptively** at each assessment for the investigational arm
  - Denominator = number of patients on treatment at time *t*
  - Handling intercurrent events:
    - Patient dropped from analysis population after progression, treatment discontinuation, or death


Communication of Results
Statistical Analysis Plan
Target Study Population
Variable (Endpoint) of Interest
Intercurrent Events
Population Level Summary
Estimand
PRO Research Objective

Communication of Results

Among patients who received one dose of drug and completed a baseline PF assessment, what is the proportion of on-treatment patients who at least maintained their physical functioning at every assessment?

	3 months	6 months	12 months	18 months
PF worsening	35 (10%)	20 (10%)	6 (10%)	2 (10%)
PF improvement/maintenance*	280 (80%)	154 (77%)	47 (78%)	13 (65%)
Missing PF assessment	35 (10%)	26 (13%)	7 (12%)	5 (25%)
Total patients on treatment	350	200	60	20
N = 500** **eligible patients + received one dose of drug + completed baseline PRO assessment				

Fabricated data *Based on pre-defined clinically relevant within-patient threshold for improvement/maintenance

Summary of Where Discussion Started
Research Objective 2: Describing Patient Perspective

Estimand attributes	Decisions to better define research objectives
Target population	One dose of drug + completed baseline PF assessment + on treatment
Variable of interest	Patients who maintained/improved PF based on pre-specified criteria at every assessment point until end of treatment
Handling of intercurrent event	
Death, disease progression, treatment discontinuation	Patient dropped from denominator after intercurrent event occurs
Population level summary	Proportion of on-treatment patients who maintained/improved PF

These case studies are **not an endorsement** of a singular study design, outcome, analysis, or visualization; rather it's meant to demonstrate how FDA may perceive physical function data in oncology

Panel Discussion # 1

What is the more appropriate or informative way of describing proportion of patients who at least maintained their PF for this scenario?

Table 1
Denominator: Total patients on treatment at time *t*

	3 months	6 months	12 months	18 months
PF worsening	35 (10%)	20 (10%)	6 (10%)	2 (10%)
PF improvement/maintenance	280 (80%)	154 (77%)	47 (78%)	13 (65%)
Missing PF assessment	35 (10%)	26 (13%)	7 (12%)	5 (25%)
Total patients on treatment	350	200	60	20
N = 500* *eligible patients + received one dose of drug + completed baseline PRO assessment				

Fabricated data

	3 months	6 months	12 months	18 months
PF worsening	35 (7%)	20 (4%)	6 (1%)	2 (0.4%)
PF improvement/maintenance	280 (56%)	154 (31%)	47 (9%)	13 (2.6%)
Missing PF assessment	35 (7%)	26 (5%)	7 (1%)	5 (1%)
Discontinued treatment	150 (30%)	300 (60%)	440 (88%)	480 (96%)
N = 500* *eligible patients + received one dose of drug + completed baseline PRO assessment				

Table 2
Denominator: PRO analysis population (N = 500)

Panel Discussion # 2

Did these findings address what you'd like to know about patient experience on the drug?

What other information are you looking for to gain more insight about patients' experience on the drug?

Questions From the Audience

- **Co-Moderators**
  - Mallorie Fiero – FDA statistician
  - Chana Weinstock – FDA clinician
  - Madeline Pe – SISAQOL

- **Panelists**
  - Andrea Ferris – Patient advocate
  - Sigrid Klaar – European regulatory and payer perspective
  - Alicyn Campbell – Industry
  - Surya Singh – Domestic payer
  - David Cella – Academic psychometrician
  - Kim Cocks – Academic statistician

Additional Panel Discussion Questions

1. We have seen how we defined estimands to describe the patient perspective. Is this feasible? What do you foresee as real-life challenges when defining PRO research objectives in this way?

2. To respond to this research objective, we defined a responder using a “cut-off” score. What are your thoughts about dichotomizing a continuous variable into patients who maintained/improved and those who did not?

Concluding Remarks

What is the key element of the estimand discussion? Did you feel a shift in your own perspective after the discussions?

Take Home Messages

- There is a need for more **well-defined research objectives** that can be matched with appropriate statistical methods
  - **Estimand framework** is an organized approach to construct a well-defined endpoint

- Lack of superiority (e.g., $p > 0.05$) **does not mean** equivalence

- There is **no one best way** to evaluate patient experience, but **standard principles and analyses** must be developed

Acknowledgements

- Raji Sridhara
- Laura Lee Johnson
- Paul Kluetz
- Bellinda King-Kallimanis
- Nirosha Lederer

BACKUP

Considerations for Addressing Intercurrent Events

Handling intercurrent events (examples)	
• Value for variable used regardless of whether or not intercurrent event occurs • Make intercurrent event part of composite endpoint	• Value for variable used until intercurrent event occurs • Restrict population of interest to subset of patients in which intercurrent event would not have happened

Analysis Plan

	Draw conclusions on treatment efficacy / clinical benefit (Confirmatory Objective)		Describe patient experience (Exploratory / Descriptive Objective)
Within-treatment arms assumption *(longitudinal design: applies to both short-term and long-term)*	**Between treatment arms objective**		
	Superiority	**Equivalence / Non-inferiority**	
1. Improvement			
a. Time to improvement	- Statistical method	Statistical method	
b. Proportion of patients with improvement at time t	- Statistical method	Statistical method	
c. Magnitude of improvement at time t	- Statistical method	Statistical method	
2. Maintenance			
a. Time to (end of) maintenance	- Statistical method	Statistical method	
b. Proportion of patients with maintenance at time t	- Statistical method	Statistical method	
c. Magnitude of maintenance at time t	- Not applicable		
3. Worsening			
a. Time to worsening	- Statistical method	Statistical method	
b. Proportion of patients with worsening at time t	- Statistical method	Statistical method	
c. Magnitude of worsening at time t	- Statistical method	Statistical method	
4. Overall effects			
a. Overall PRO score over time	- Statistical method	Statistical method	
b. Response patterns / profiles	- Statistical method	Statistical method	

Slides provided by SISAQOL Consortium