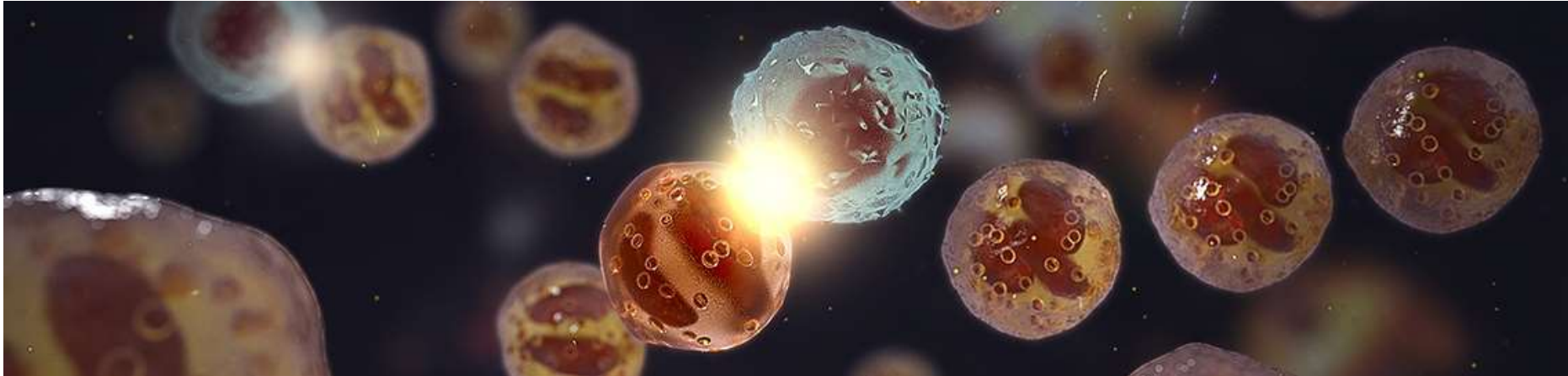# Estimands in Chronic Symptomatic Diseases

Martin Jenkins

Statistical Science Director, AstraZeneca, Cambridge

September 2017

Symptomatic Diseases

These are my views and do not necessarily represent the views of AstraZeneca

Today's presentation will draw on experience from the disease areas below.
- Rheumatology
  - **Rheumatoid Arthritis (RA)**
  - **Systemic Lupus Erythematosus (SLE)**
  - **Psoriatic Arthritis (PsA)**
- Dermatology
  - **Psoriasis**
  - **Atopic Dermatitis (AD)**
- Respiratory
  - **Asthma**
  - **Chronic Obstructive Pulmonary Disease (COPD)**

Although much of this is my direct experience I also draw on work by others within AstraZeneca for some examples. With thanks to: Lesley France, David Wright, Anna Berglind, Chris Miller, Patrick Darken, Mattis Gottlow

Contents

- Common characteristics of symptomatic diseases

- Estimand choice for symptomatic diseases
  - Attributable estimands

- Example scenarios
  - From automimmune, rheumatology, respiratory settings

- Some asides
  - Composite endpoints, Randomised withdrawal designs etc

- Summarising general considerations in estimand choice
  - Decision tree

Common characteristics of Symptomatic Diseases

Endpoints - Symptoms

- Chronic diseases where symptoms need to be managed long term
- Endpoints reflect the extent or severity of symptoms
- May be directly measured or patient / physician reported
- Symptomatic improvements (of short-acting treatments) may be reversible
- Composite endpoints may be used to reflect impact on several body systems
- Diseases are relapsing and patients experience a form of flare / exacerbation
- Not (yet) talking about cure, rarely even about disease modification

> Primarily interested in the effect of just the initially randomised treatment on the control of symptoms

Common characteristics of Symptomatic Diseases

Intercurrent events - Rescue medications

- In some diseases there are reasonably well-defined treatment pathways

- Often background or rescue treatments as part of treatment regimen

- Switching or rescue expected in some settings, to move patients off ineffective treatments if they do not respond

- Treatment escalation may come with additional burden and is not necessarily desirable, eg high dose corticosteroids or off-label treatments

Need to consider the "intercurrent event" of rescue treatment

Recap - ICH E9 (R1) – Estimand terminology

- The ICH E9 Addendum proposes at least 5 strategies for handling "intercurrent events"
- In our scenario these include treatment escalation, rescue, discontinuation, withdrawal, death…

Estimand	Handling of intercurrent event
"Treatment policy"	Actual values of the variable regardless of whether the intercurrent event has occurred.
"Composite"	Modified definition of the variable or the summary measure such that an intercurrent event becomes a component of the outcome.
"Hypothetical"	Values of the variable under some hypothetical conditions where an intercurrent event would not happen.
"Principle stratum"	Restrict population of interest to the stratum of patients in which an intercurrent event would not have happened.
"While on treatment"	Values of the variable up to the time of the intercurrent event, rather than at a planned assessment time point.

Estimands for symptomatic endpoints

- Generally aim to seek *inference about the initially randomised treatment* in controlling symptoms

- Rarely the aim to assess *treatment policies* because symptomatic measures may respond quickly to changes in treatment and so not reflect the effect of the therapy of interest.

- Benefits of symptom reduction may be lost on the discontinuation of treatment and commonly patients with poor initial response are "rescued" with additional treatments.

> Imagine a study of a short-acting symptom control treatment vs placebo
>
> All patients do poorly in the control arm and are rescued to alternative treatment, where they all respond. Would a treatment policy comparison reflect the benefits of the randomised treatment?

- Despite this, on occasion treatment policy estimands can appear to be a blanket policy for some agencies.

Estimands for symptomatic endpoints

- O'Neill and Temple (2012): *"In symptomatic settings, it is not the usual practice to continue to assess effectiveness in subjects after they have stopped taking the assigned treatment (ITT approach), as the drug's effect is assumed to be lost; also, in many cases, an alternative drug is started, and this could influence the outcome for a subject."*

- ICH E9, R1, (2017): *"The goal of a treatment may be control of clinical signs or symptoms in a disease area where multiple alternative treatments exist… The specification of how to account for intercurrent events might be based on understanding the treatment effect if the alternative treatment was not available… In some circumstances, answers to these questions might be more relevant than e.g. the quantification of the effects of a treatment policy that does not distinguish whether or not a patient has taken an alternative treatment."*

- Holzhauer, Akacha, Bermann (2015), discuss the scenario in diabetes: *"…the scientific question of interest in … is whether and by how much an intervention would have differed from a comparator in terms of <<average-achieved glucose levels assessed by HbA1c>> after a sufficiently long period of time, if no rescue medication had been initiated. Our problem is simply that the outcome of interest cannot be directly observed for all patients, because withholding rescue medication would be unethical…"*

8 **O'Neill RT, Temple R.** The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. Clin Pharmacol Ther. 2012;91(3):550-4.
Holzhauer B, Akacha M, Bermann G. Choice of estimand and analysis methods in diabetes trials with rescue medication. Pharm Stat.2015; 14(6): 433-447

Estimands for symptomatic endpoints

- So the scientific question of interest often relates to a scenario where rescue had not been mandated

- Entirely counterfactual (or hypothetical) comparison that could not be observed in practice on ethical grounds.

- Previously, the literature considered several such hypothetical families, eg:

Estimand	Hypothetical scenario considered
"Efficacy" (de jure)	The difference if all patients had continued on their initially randomised treatment without treatment escalation or rescue
"Effectiveness" (de facto)	The difference if all patients had continued in the study without treatment escalation or rescue, accounting for different hypothetical outcomes depending upon rescue / discontinuation / withdrawal.

Holzhauer B, Akacha M, Bermann G. Choice of estimand and analysis methods in diabetes trials with rescue medication. Pharm Stat.2015; 14(6): 433-447
Phillips et al. Estimands: discussion points from the PSI estimands and sensitivity expert group. Pharm Stat. 2017; 16 (1): 6-11
Mallinckrodt CH, Lin Q, Lipkovich I, Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. Pharm Stat. 2012;11(6):456-61
Leuchs AK, Zinserling J, Brandt A, Wirtz D, Benda N. Choosing Appropriate Estimands in Clinical Trials. Ther Innov Regul Sci. 2015;49(4):584-92

Addressing the Attributable Estimand

- Often we refer to *"the effect attributable to initially randomised treatment"*

- Implies data after rescue or the initiation of prohibited medications should not be used
- Assume patients who required rescue would have done worse than the completers <u>had rescue not been available</u>.
- Leaves open what the method of imputation/estimation should be?

- If there is a "**rescue rule**" and models include all endpoints that imply a need for rescue then a case could be made that this is MAR as there is full information to predict rescue.

- If we are interested in data that cannot be observed (a counterfactual scenario) then this typically requires untestable assumptions to be made.
- This is when the robustness of the results must be demonstrated using sensitivity analyses.

Phillips et al. Estimands: discussion points from the PSI estimands and sensitivity expert group. Pharm Stat. 2017; 16 (1): 6-11
Mallinckrodt CH, Lin Q, Lipkovich I, Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. Pharm Stat. 2012;11(6):456-61
Leuchs AK, Zinserling J, Brandt A, Wirtz D, Benda N. Choosing Appropriate Estimands in Clinical Trials. Ther Innov Regul Sci. 2015;49(4):584-92

Handling non-treatment related intercurrent events

- If assessing the difference "attributed to treatment" then how should we handle intercurrent events that are not treatment related?
- A reasonable question could be to ask what difference is attributable to treatment if no such events occurred, but whilst accounting for the adverse impacts of treatment as well as the potential benefits:

Intercurrent event	"Efficacy"	"Effectiveness"	"Attributable"
Rescued or initiated prohibited medications	Hypothetical outcome - stayed on randomised treatment	Hypothetical outcome - stayed on randomised treatment	Hypothetical outcome - stayed in study without treatment escalation
Non-treatment related discontinuation *(eg AE not reasonably treatment related)*	Hypothetical outcome - stayed on randomised treatment	Hypothetical outcome - stayed in study without treatment escalation*	Hypothetical outcome - stayed on randomised treatment
Treatment related discontinuation *(eg lack of efficacy or treatment-related AE)*	Hypothetical outcome - stayed on randomised treatment	Hypothetical outcome - stayed in study without treatment escalation*	Hypothetical outcome - stayed in study without treatment escalation*

*Note: - Assumes there is not continued follow-up post discontinuation. If there is observed data after discontinuation, but without rescue then this may be used here.
In all cases these are contingent upon the given covariates.

Addressing the Attributable Estimand

- For these scenarios some form of MNAR imputation is required, often multiple imputation based.

- Many possibilities exist, for imputing some sort of "worse outcome" including:

  - **Delta adjustment** (reduce the effect by a fixed increment)

  - **Imputation using a reference arm** (variations on "jump to reference" etc)

  - **Rank-based imputation** (some form of "worse case" type scenario)

  - **Imputation using a percentile or quantile** (eg, centre on X^{th} percentile)

  - **Return to baseline** (centre around this, rather than single imputation BOCF)

- Tipping point analyses could be performed as sensitivity analyses to vary the parameters used.
- Some overlap here between "hypothetical" or "composite" approach to intercurrent events

[12] **Holzhauer B, Akacha M, Bermann G.** Choice of estimand and analysis methods in diabetes trials with rescue medication. Pharm Stat.2015; 14(6): 433-447
Mehrotra DV, Liu F, Permutt T. Missing data in clinical trials: control-based mean imputation and sensitivity analysis. Pharm stat. 2017;

Considerations in estimand choice

- So many possibilities, even aside from treatment policy estimands. No universal all-purpose solution.

- ICH E9 (R1): *"The set of intercurrent events for consideration will depend on the specific therapeutic setting and trial objective." "The relevance of each strategy will depend on the therapeutic and experimental context."*

- Several aspects be considered when assessing the relevant estimand for a clinical trial:
    - What is the treatment pathway and context?
    - Is the endpoint direct, symptomatic and reversible?
    - Interested in maximum efficacy or effectiveness in practice?
    - Is the effect of treatment retained after discontinuation?
    - Does treatment discontinuation imply a poor outcome?

- ***Let's consider some examples…***

Framework for Examples

Endpoint
Drug type
Treatment pathway

Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
Shorthand term	*Specific description:* • *Population* • *Variable* • *Timeframe* • *Handling of intercurrent events* • *Summary measure…*	*Yes/no -* *This is important in determining study requirements*	*What we want to impute when addressing this. Not part of the estimand as such.*	*Statistical model that could be used to achieve this*

Systemic Lupus Erythematosus

- Chronic, systemic, autoimmune disease affecting many organs.

- **Treatments:**

  - Milder disease: NSAIDs, low-dose corticosteroids, hydroxychloroquine

  - More severe disease: high dose steroids, immunosuppressants, B-cell modulators

- Achieving disease control through the long-term use of high dose steroids or cytotoxic drugs may not be desirable. These variously come with tolerability concerns or lack of proven efficacy.

- **Endpoints:** (eg BILAG grade) are defined as symptoms requiring the need for immunosuppresives or increase in corticosteroids

Most common symptoms of
Systemic lupus erythematosus

Systemic:
- Low-grade fever
- Photosensitivity

Psychological
- Fatigue
- Loss of appetite

Mouth and nose
- Ulcers

Face
- Butterfly rash

Muscles
- Aches

Pleura
- Inflammation

Joints
- Arthritis

Pericardium
- Inflammation

Fingers and toes
- Poor circulation

Example estimand - SLE

Scenario 1

Endpoint: SRI-4, binary composite response endpoint
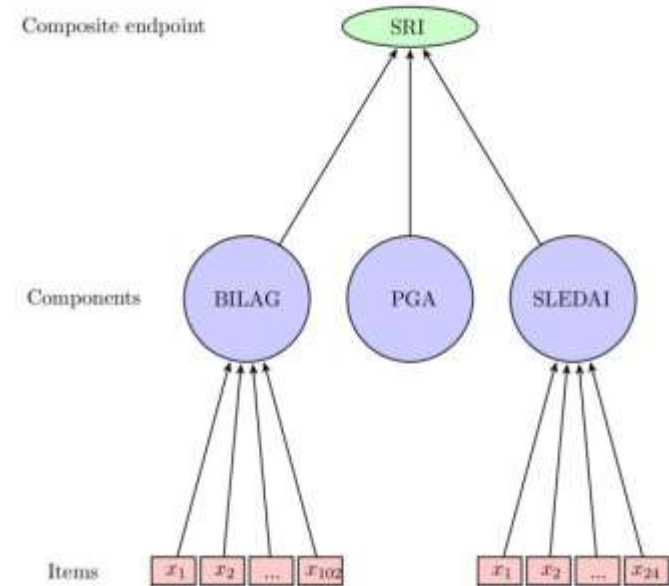**Drug type**: Biologic
Treatment pathway: NSAIDs, steroids, topical treatments, belimumab, systemic immunomodulators or anti-malarials,

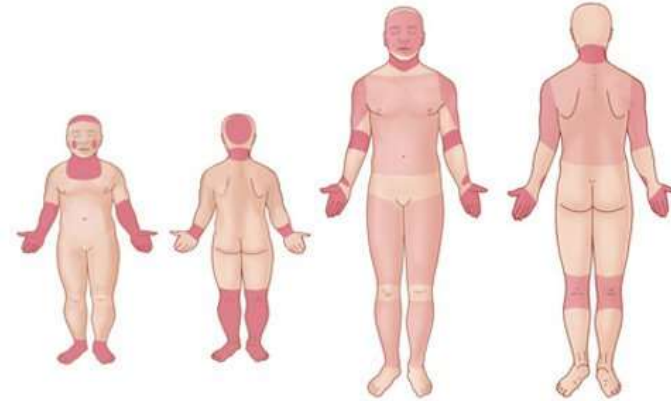| Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
"Composite responder"	"Difference between treatments in proportion of SRI-4 responders at Week 52 where subjects who **start restricted medication or discontinue IP** are considered treatment failures."	**No**, only data up until discontinuation is required to define endpoint	Subjects who start restricted medication or discontinue IP are defined as non-responders.	Endpoint definition effectively gives complete data. Analyse using Cochran-Mantel-Haenszel method

Aside – Composite endpoints

- "Responder analyses" are an attractive solution

- Well established in rheumatology and accepted by regulators
- Familiar and understandable to clinicians in these areas

- Some downsides
  - Dichotomisation, reducing statistical information
  - Require sensitivity analyses - eg cumulative distribution
  - Equivalent analyses on a continuous scale are difficult

- Multi-level nature of missing data at the component or item level

- Joint modelling of continuous components and binary items / withdrawals could be considered (Wason and Jenkins 2016)

17 **Wason JMS, Jenkins M.** Improving the power of clinical trials of rheumatoid arthritis by using data on continuous scales when analysing response rates: an application of the augmented binary method. Rheumatology. 2016; 55(10):1796-1802.

Atopic dermatitis

- Chronic, inflammatory skin disease characterized by itchy, red, swollen, cracked skin.

- **Treatments:**

  - Milder disease: topical treatments – Emoliants, Topical corticosteroids (TCS)

  - Severe disease: few options until recently. Biologics just entering the market.

- Treatment escalation may not be permanent, but could just be to control disease. Benefits could persistent after discontinuation.

- **Endpoints:** include investigator or patient defined response scales which can be dichotomized (eg IGA, EASI-75)

Common Sites of Eczema Outbreaks

Example estimand – Atopic dermatitis

Scenario 2

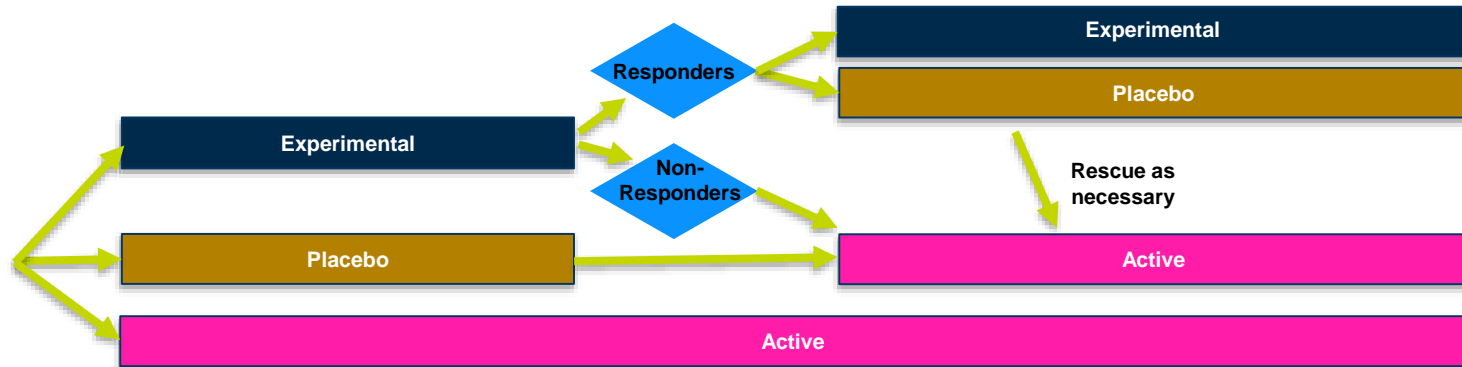*Endpoint*: IGA, binary response endpoint
Drug type: Biologic
Treatment pathway: Emollients, topical steroids / topical calcineurin inhibitors, immunosuppresants, oral/systemic steroids

Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
"Composite responder"	"Difference between treatments in proportion of IGA responders at Week 16 where subjects who **start restricted medication or withdraw from study** are considered treatment failures."	**Yes**, uses data up until the start of restricted medications or study end, regardless of discontinuation of randomised treatment	Subjects who start restricted medication or **withdraw from study** are counted as non-responders.	Imputation effectively gives complete data. Analyse using Cochran-Mantel-Haenszel method

Aside - Randomised withdrawal designs

- In dermatology (eg psoriasis) randomised withdrawal designs are often used to study maintenance
- Patients who respond to initial treatment are re-randomised to stay on treatment or switch to placebo
- As such, estimands here must specify the population of re-randomised individuals as well as the intercurrent events of re-treatment.

- Eg: "*Amongst patients with IGA response at Week 16*, the difference between treatments in proportion of IGA responders at Week 52 where subjects who receive rescue medication (retreatment) or withdraw from the study are considered treatment failures."
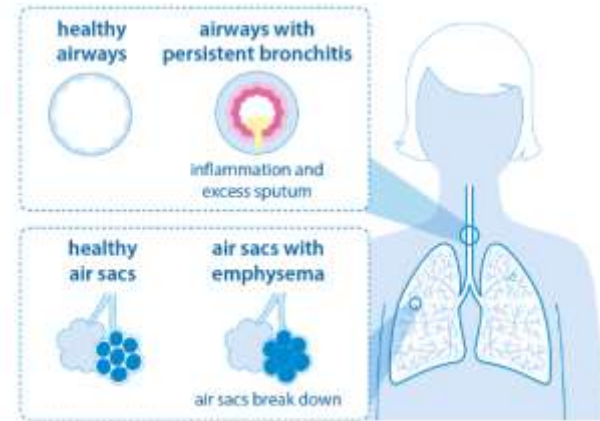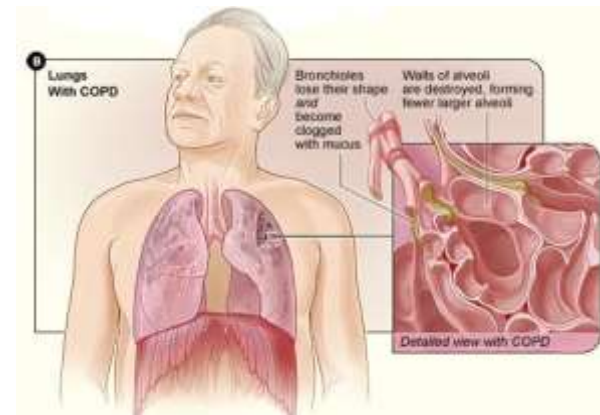
Aside - Can we ever consider MCAR scenarios?

- Scenario - study is discontinued prematurely for operational reasons independent of the efficacy findings to date. Some patients have had chance to complete the study, others have not. (Primary endpoint is binary outcome at 16 weeks)

- Question of interest? "Difference in proportion of responders between treatments where discontinuation is considered treatment failure, ***had all patients had the opportunity to complete the study***."

- "**Principle strata**" / "**responder**" analysis: - define an analysis set of those patients who were randomised early enough to have <u>the opportunity</u> to complete the study (ie 16 weeks before closure). Apply non-responder imputation

- Hypothetical "**efficacy**" / "**responder**" analysis: - use a pattern mixture approach where missing data due purely to premature study closure is considered MCAR. Otherwise use non-responder imputation.

Chronic Obstructive Pulmonary Disease

- Encapsulates a group of long term lung conditions, including bronchitis and emphysema.

- **Treatments:**

    - Relievers: short-acting medications in the case of shortness of breath. Generally a background medication in clinical trials.

    - Maintenance: various combinations of inhaled corticosteroids, LAMAs and LABAs are available (or in development).

- **Endpoints:** include pulmonary function testing, dyspnea or symptom scores and exacerbations.

- Some long term maintenance treatments are aimed at improving lung function and managing symptoms with others are more aimed at preventing exacerbation.

Example estimand – COPD

Scenario 3

Endpoint: - Forced Expiratory Volume in 1 second (FEV1) at morning pre-dose trough
Drug type: - Inhaled maintenance treatment
Treatment pathway: Short-acting rescue inhaler, combination bronchodilator treatments, ICS containing maintenance

Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
Hypothetical "Efficacy"	"Difference between treatments in mean change from baseline in trough FEV1, at Week 24 **had all subjects remained on randomised treatment**."	No, only use data up until discontinuation	Subjects who discontinue IP have remaining data imputed* to reflect the expected outcome had the subject remained on the study treatment without using restricted medications (assume MAR)	MMRM model under MAR is applied to the data prior to treatment discontinuation.

*Note, no actual imputed value per patient, mixed model used to derive result at treatment group level

Example estimand – COPD

Scenario 4

Endpoint: - Forced Expiratory Volume in 1 second (FEV1) at morning pre-dose trough
Drug type: - Inhaled maintenance treatment
Treatment pathway: Short-acting rescue inhaler, combination bronchodilator treatments, ICS containing maintenance

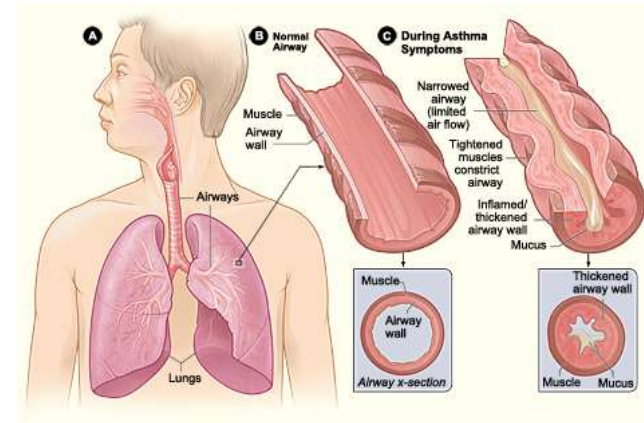| Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
Hypothetical "Attributable"	"Difference in mean change from baseline in trough FEV1 at Week 24 **that could be attributed to treatment had all subjects remained on the study without access to restricted medications."**	No, only use data up until discontinuation or use of prohibited medications	Discontinue for treatment related reasons or start restricted meds = imputed to reflect **expected outcome had subject remained in study without ability to escalate treatment.** Non-treatment related reasons = **impute as if subject remained on treatment.**	Data after prohibited meds set to missing, then apply MI model with different patterns based on reason for missingness. **MAR for those where unrelated. Where potentially related, impute poorer outcome** centred at estimate of 5th percentile of control arm.

Disease Exacerbations

- **Symptomatic endpoints** = most interested in hypothetical estimands…
- **Irreversible disease modification or morbidity** = suit a treatment policy estimand…

- Disease **exacerbations** (or flares) exhibit features of both of these.
  - Acute events of a limited duration for which the patient can often be treated and returned to a status of symptom control
  - But some exacerbations can have serious outcomes
  - Patients with a history of exacerbations are much more likely to exacerbate again
  - Treatment strategies should take exacerbation risk into consideration

- So less clear cut whether a "treatment policy" approach is justified
- Some health authorities tend to prefer this as the primary approach and to require continued data collection after discontinuation of randomised treatment.

- There should be consideration of estimands that do not treat post-discontinuation data as MAR.

Asthma

- Inflammatory disease of the respiratory airways, triggered by allergens and other autoimmune mechanisms

- **Treatments:**

  - Relievers: inhalers with short-acting drugs can be used for short term relief as required

  - Maintenance: inhaled corticosteroids (and LABAs) are used to control symptoms over the longer term.

- Biologics are just becoming available as an option for patients with severe asthma driven by eosinophils.

- **Endpoints:** include pulmonary function, symptom scales and exacerbation rates

Example estimand – Severe Asthma

Scenario 5

Endpoint: - Annualised Asthma Exacerbation rate
Drug type: - Biologic maintenance treatment
Treatment pathway: Short-acting rescue inhaler, ICS containing maintenance, Biologic is added on top of existing inhalers

Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method
"Treatment policy"	"Ratio of exacerbation rates between treatments over 52 weeks, *regardless of treatment received*."	*Yes*, data collected to end of study regardless of treatment discontinuation	Ideally seek complete follow-up. Subjects who withdraw from the study have their event rate for the remaining period imputed to reflect the expected counts had the subject remained in the study (ie conditioning on the events observed)	Negative binomial generalised linear model with offset for log(follow-up time). If near-complete follow-up then may assume MAR.

Example estimand – Severe Asthma

Scenario 6

Endpoint: - Annualised Asthma Exacerbation rate
Drug type: - Biologic maintenance treatment
Treatment pathway: Short-acting rescue inhaler, ICS containing maintenance, Biologic is added on top of existing inhalers

Estimand label	Estimand Description	Uses Post-discontinuation data?	Imputation	Possible Analysis Method (Estimator)
Hypothetical "Effectiveness"	"Ratio of exacerbation rates between treatments over 52 weeks, **had all subjects remained in the study**."	**No**, only use data up until discontinuation	Subjects who discontinue have their event rate for the remaining period imputed to reflect the expected counts had the subject remained in the study, **accounting for an assumed loss of efficacy post-discontinuation**.	Multiple imputation model with different patterns whether reason for missingness is treatment related. MAR for those where unrelated. Where potentially related, **impute event rate based upon control arm.**

Further considerations

In several areas, the current thinking on estimands could be further developed

- **Non-inferiority**: - Previously the thinking has been driven more by what is or is not "conservative" rather than relevant questions / estimands. Is there still a role for per protocol analyses?

- **Disease modification**: - Hope in the future this could become a realistic target. What estimands will then be relevant and what designs are needed? More likely to be treatment policy analyses

- **Long term extensions**: - Expected in many chronic diseases, but in the past have often been uncontrolled or analysed on an "as observed" basis. How can sensible estimands be realistically assessed?

- **Crossover studies**: - What is the relevant vocabulary for crossover estimands? Those who tolerate either treatment sufficiently to complete both periods defined as a principal stratum?
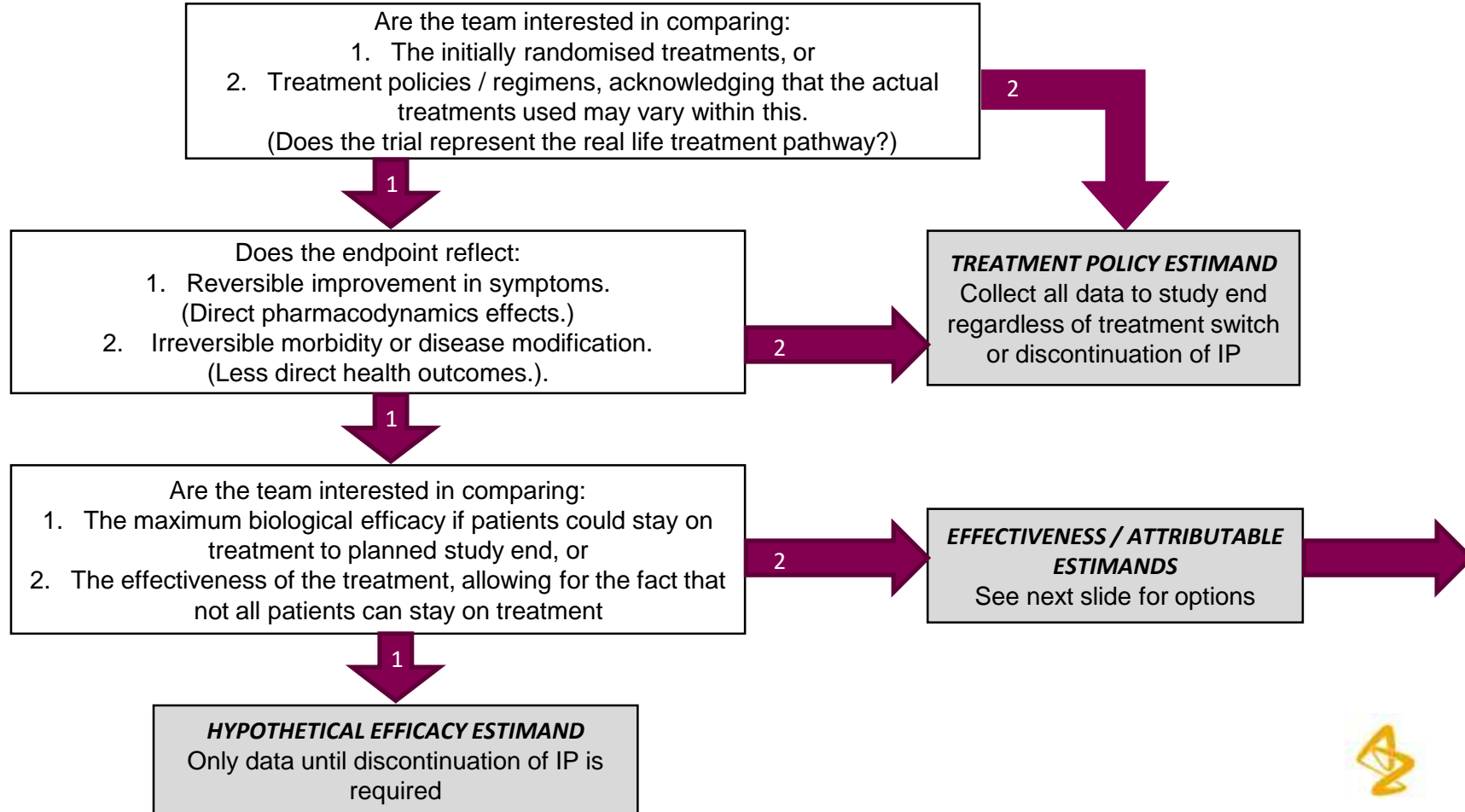
General rules of thumb

- So there is no general one-size-fits all conclusion
- But what could be helpful? Some general rules of thumb in estimand choice?

- As described in these examples, some aspects to consider would be:
    - What is the treatment pathway and context?
    - Is the endpoint direct, symptomatic and reversible?
    - Interested in maximum efficacy or effectiveness in practice?
    - Is the effect of treatment retained after discontinuation?
    - Does treatment discontinuation imply a poor outcome?

- Can we put some structure on this thinking to identify common scenarios? See the following *flow chart* for an attempt.

Simplified decision tree

Are the team interested in comparing:
1. The initially randomised treatments, or
2. Treatment policies / regimens, acknowledging that the actual treatments used may vary within this.
(Does the trial represent the real life treatment pathway?)

2

1

Does the endpoint reflect:
1. Reversible improvement in symptoms. (Direct pharmacodynamics effects.)
2. Irreversible morbidity or disease modification. (Less direct health outcomes.).

2

TREATMENT POLICY ESTIMAND
Collect all data to study end regardless of treatment switch or discontinuation of IP

1

Are the team interested in comparing:
1. The maximum biological efficacy if patients could stay on treatment to planned study end, or
2. The effectiveness of the treatment, allowing for the fact that not all patients can stay on treatment

2

EFFECTIVENESS / ATTRIBUTABLE ESTIMANDS
See next slide for options

1

HYPOTHETICAL EFFICACY ESTIMAND
Only data until discontinuation of IP is required

EFFECTIVENESS / ATTRIBUTABLE ESTIMANDS

Could effects on the endpoint after treatment discontinuation, but before new treatments are initiated be attributed to the randomised treatment?
1. Yes – effects could be retained for a period of time before new meds are initiated
2. No, effects don't persist for long, new medications expected to be started quickly

1

How should the initiation of alternative medications be interpreted?
1. See this as a failure of treatment
2. Wish to impute what would have happened if these medications weren't taken

2

How should discontinuation or the initiation of alternative medications be interpreted?
1. See this as a failure of treatment
2. Wish to impute what would have happened if these medications weren't taken

1

RESPONDER ESTIMAND
Collect data until initiation of new treatment regardless of discontinuation of IP

2

HYPOTHETICAL EFFECTIVENESS ESTIMAND
Collect data until initiation of new treatment regardless of discontinuation of IP

1

RESPONDER ESTIMAND
Only data until discontinuation of IP is required

2

HYPOTHETICAL EFFECTIVENESS ESTIMAND
Only data until discontinuation of IP is required

Conclusions

- Many chronic symptomatic diseases have similar considerations around estimands

- No single solution, but common principles to consider when defining the relevant estimands for a study

- Treatment policy estimands are unlikely to be the focus for such scenarios, although these analyses apply less assumptions than with hypothetical estimands and can be attractive to some regulators

- Possible tension between most relevant question vs most robust analysis?

- Responder analyses offer an option, but invariably "hypothetical" effectiveness estimands will be of relevance as we look to establish the differences attributable to treatment

- Sponsors (statisticians) should continue to propose the use of relevant estimands in health authority correspondence.

Thank you

Confidentiality Notice