Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Contemporary Clinical Trials 108 (2021) 106494

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Why ITT analysis is not always the answer for estimating treatment effects in clinical trials

Oliver N. Keene [a,*], David Wright [b], Alan Phillips [c], Melanie Wright [d]

[a] Biostatistics, GlaxoSmithKline Research and Development, Brentford, Middlesex, UK
[b] Data Science & Artificial Intelligence, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK
[c] Biostatistics, ICON Clinical Research UK Ltd, Marlow, Buckinghamshire, UK
[d] Clinical Development and Analytics, Novartis Pharma AG, Basel, Switzerland

ARTICLE INFO

Keywords:
Estimands
Estimation
ITT
Intent-to-Treat
Treatment policy
Missing data
Multiple imputation

ABSTRACT

For many years there has been a consensus among the Clinical Research community that ITT analysis represents the correct approach for the vast majority of trials. Recent worldwide regulatory guidance for pharmaceutical industry trials has allowed discussion of alternatives to the ITT approach to analysis; different treatment effects can be considered which may be more clinically meaningful and more relevant to patients and prescribers.

The key concept is of a trial "estimand", a precise description of the estimated treatment effect. The strategy chosen to account for patients who discontinue treatment or take alternative medications which are not part of the randomised treatment regimen are important determinants of this treatment effect. One strategy to account for these events is treatment policy, which corresponds to an ITT approach. Alternative equally valid strategies address what the treatment effect is if the patient actually takes the treatment or does not use specific alternative medication. There is no single right answer to which strategy is most appropriate, the solution depends on the key clinical question of interest.

The estimands framework discussed in the new guidance has been particularly useful in the context of the current COVID-19 pandemic and has clarified what choices are available to account for the impact of COVID-19 on clinical trials. Specifically, an ITT approach addresses a treatment effect that may not be generalisable beyond the current pandemic.

1. Introduction

Many statistical analyses of clinical trials are described simply as following the ITT principle. The ITT principle has sometimes been interpreted as referring only to the population to be studied (i.e. all randomised patients) but actually requires follow-up of all patients to the end of the planned study period and inclusion of all available data regardless of whether the patient completed randomised treatment or took alternative medication [1]. This strict interpretation, however, does not account for the diversity in the patient journeys since the start of the trial [2] and the potential impact that this may have on the interpretation of the findings. Per protocol analysis has sometimes been proposed as an alternative to ITT analysis; the role of Per Protocol analysis is discussed in section 4.2 below.

2. Example trials in Diabetes and COPD

To illustrate the issue, consider the following trial in type II diabetes mellitus where dapagliflozin was compared to placebo [3]. The primary endpoint was change in HbA1c from baseline to 24 weeks. Use of placebo was only considered ethical if patients who experienced lack of efficacy could receive rescue medication, and collection of HbA1c continued until the end of the trial, irrespective of whether rescue medication was received or not. The published analysis was performed using analysis of covariance with the baseline value as a covariate; the analysis excluded HbA1c data after initiation of rescue medication.

However, such an analysis does not conform to the ITT principle of inclusion of all available data regardless of whether the patient complied with the requirements of the protocol. When the trial was submitted as evidence of efficacy for potential regulatory approval, the FDA reviewer at the licensing authority performed an alternative analysis using all

* Corresponding author at: GlaxoSmithKline Research and Development, 980 Great West Road, Brentford, Middlesex TW8 9GS, UK.
E-mail address: oliverkeene405@gmail.com (O.N. Keene).

https://doi.org/10.1016/j.cct.2021.106494
Received 22 February 2021; Received in revised form 25 May 2021; Accepted 24 June 2021
Available online 26 June 2021
1551-7144/© 2021 Published by Elsevier Inc.

available data regardless of rescue [4,5].

This choice made in the statistical analysis had an impact on the corresponding magnitude of the effect (Table 1); the treatment benefit on HbA1c for dapagliflozin 10 mg compared to placebo in the absence of rescue medication was 0.66%, but the benefit regardless of use of rescue was 0.45% [4]. The lower treatment effect for dapagliflozin from the ITT analysis called into question the benefit of the medication as it appeared to indicate a lesser treatment effect than other approved agents at that time [6], although it is not completely clear what analysis strategy was used for these agents.

This is not actually a statistical analysis issue; the disagreement is more fundamental. There are two different implied scientific questions which are being addressed:

a) "what is the treatment effect if no rescue medication had been used?"
b) "what is the treatment effect whether or not rescue medication is used?".

There is no right or wrong answer here, as one approach addresses the direct efficacy of dapagliflozin while the other approach assesses the policy of starting treatment with dapagliflozin with or without the use of rescue medication as required. Given these fundamentally different questions, it is not surprising that there were different estimates. This illustrates the importance of clarity on the clinical questions where are being addressed by the estimate.

The treatment effects reported in trial publications are often far from clear. For example, 'compare dapagliflozin to placebo based on the HbA1c at 24 weeks in all randomised patients' is too vague as it does not allow for the use of rescue treatment which can clearly impact the results. The treatment effect may be defined by the description of the statistical analysis but this is not transparent.

Another example of where ITT analysis can lead to issues in interpretation is the OPTIMAL trial [7]. This was a year-long study in COPD, in which all patients received tiotropium. Patients were randomised to receive three add-on treatments: placebo, salmeterol or the combination of salmeterol and fluticasone. The OPTIMAL study planned to obtain information on exacerbations from subjects withdrawing from treatment.

Of 175 patients who stopped their randomised treatment, 110 (63%) provided off-treatment data [8]. Withdrawals from treatment were fewer in the arm that received the combination of salmeterol and fluticasone compared to the two other arms (Table 2). Because salmeterol and fluticasone were available as marketed products, patients stopping their randomised treatment could receive these medications or ones in the same class (salmeterol is a long acting β_2-agonist and fluticasone is an inhaled steroid). Data on concurrent medications obtained after discontinuation of randomised treatment showed that in the tiotropium plus placebo arm, 70% of withdrawn patients received an open-label long-acting β_2-agonist and inhaled steroid combination inhaler for the remainder of the study [7], i.e. 70% of withdrawn patients in the placebo add-on treatment group switched to the same or similar medication to the treatment group receiving salmeterol and fluticasone combination.

In a separate secondary publication [8], the OPTIMAL investigators contrast the ITT analysis with an on-treatment analysis for the secondary endpoint of rate of exacerbations of COPD. The estimates presented are for the relative rate of exacerbations for fluticasone–salmeterol as add-on compared to placebo. In the ITT analysis, all data was included, while for the on-treatment analysis, data after discontinuation of randomised treatment was excluded. The magnitude of reduction in exacerbation rate for the on-treatment negative binomial model analysis is 25% (relative rate = 0.75; 95% CI: 0.55–1.01; $p = 0.06$), an appreciably larger reduction than the estimated 15% reduction from the ITT negative binomial model analysis (relative rate = 0.85; 95% CI: 0.65–1.11; $p = 0.23$).

Pharmaceutical industry trials are governed by a series of scientific guidelines from the International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). These difficulties in expressing clearly the treatment effect to be estimated in clinical trials led to a new addendum to the ICH E9 guideline on statistical principles for clinical trials [9]. This addendum introduced the concept of an "estimand" to refer to the treatment effect to be estimated in a clinical trial. The addendum asks that the estimand be precisely defined in the trial protocol alongside the objectives of the study.

The introduction of this framework for pharmaceutical trials has influenced the statistical community to think more deeply about the connection between what is being estimated (the estimand) and how the estimate is calculated (the analysis). However, practicing clinicians remain largely disconnected; "most of the focus of the estimand discussion today is still within the statistical community" [10]. The goal of this paper is to explain the concept of estimands and to stimulate discussion on appropriate strategies for events that complicate the estimation and interpretation of treatment effects.

Table 1
Treatment comparison of dapagliflozin 10 mg to placebo in the Ferrannini et al. trial.

	Placebo	Dapagliflozin 10 mg	Difference (Dapagliflozin – Placebo)
Mean change from baseline in HbA1c			
N	75	70	
Analysis excluding data after rescue (95% CI)	−0.23%	−0.89%	−0.66% (−0.36%, −0.96%)
ITT analysis (95% CI)	−0.45%	−0.91%	−0.45% (−0.19%, −0.72%)
Patients using rescue medication, n (%)	9 (12%)	0	

Table 2
Treatment comparison of salmeterol + fluticasone + tiotropium to tiotropium alone from the OPTIMAL trial.

	Add on treatment to Tiotropium		
	Placebo	Salmeterol	Salmeterol + fluticasone propionate
N	156	148	145
Patients stopping treatment, n (%)	74(47%)	64 (43%)	37 (26%)
Relative rate of COPD Exacerbations: (salmeterol + fluticasone propionate)/placebo			
ITT analysis (95% CI)[1]		0.85 (95% CI: 0.65, 1.11; $p = 0.23$); 15% reduction	
Analysis excluding data after treatment discontinuation (95% CI)[1, 2]		0.75 (95% CI: 0.55, 1.01; $p = 0.06$); 25% reduction	

1. Negative binomial analysis.
2. Only 95% CI provided in publication; estimate obtained as mid-point of CI.

3. Estimands and intercurrent events

3.1. Definition of an estimand

The five elements of an estimand are a) treatments, b) population, c) variable (or endpoint), d) summary measure and e) strategy for each relevant intercurrent event. Four of these (treatments, population, variable and summary measure) are familiar from current trial descriptions. Treatments refers to the treatment regime being studied, including background and rescue medications, and the alternative treatment regime to which comparison will be made. The population describes the patient population of interest and typically consists of all randomised patients. The variable corresponds to the outcomes of interest for the individual patient i.e. the measurements taken on a participant in the trial and functions of these measurements. The summary measure is the group level summary that is the basis for the treatment comparison e.g. difference between means, difference in response proportions or hazard ratio.

The novel element in the definition of an estimand is pre-specification of anticipated intercurrent events together with the strategy to be used for each of these. Intercurrent events are events occurring after treatment initiation that affect either the interpretation of the outcome of interest or prevent collection of the measurements associated with the outcome of interest.

Such events include intake of rescue medication, use of medication prohibited by the protocol, treatment switching, discontinuation of randomised treatment and death (when not the endpoint of interest). Study discontinuation is not considered as an intercurrent event, instead it determines whether or not the value for the variable of interest is observed (see section 3 below on missing data). Randomisation does not help prevent intercurrent events or ensure they are equally distributed.

3.2. Strategies for intercurrent events

Table 3 shows five potential strategies for intercurrent events. Each of these results in a different estimand of interest.

It is important to note that different strategies can be applied for different intercurrent events. For example, a treatment policy strategy could be applied for discontinuation of randomised medication, and a hypothetical strategy for use of alternative medication.

A treatment policy strategy uses the actual values of the variable regardless of whether the intercurrent event has occurred. When using this strategy, it is important to ensure data are still collected after the intercurrent event as these data are included in the analysis. This approach corresponds to the ITT principle of comparing treatment policies.

For a composite strategy, the intercurrent event becomes part of the endpoint. This strategy is useful when the intercurrent event provides valuable information on its own regarding the effect of a treatment. For the dapagliflozin trial above, a patient taking rescue medication intake as required could be considered as a treatment failure. A composite strategy is commonly used as part of a binary or time-to-event endpoint, but this is not required [11].

Table 3
Potential strategies for intercurrent events.

Strategy	Relevant values for patients with the intercurrent event
Treatment Policy	Actual values of the variable regardless of whether the intercurrent event has occurred
Composite	Modified definition of the endpoint incorporating the intercurrent event
Hypothetical	Values the variable would have taken in the hypothetical scenario defined
While on-treatment/ at-risk	Values of the variable up to the time of the intercurrent event
Principal Strata	Restrict population of interest to patients who would not experience the intercurrent event

A hypothetical strategy for an intercurrent event estimates the treatment effect that would have been observed in a hypothetical scenario where no subject experienced the intercurrent event. It is important to note that, because the hypothetical outcomes needed to estimate the treatment effect are by definition not observable, the statistical analysis relies on making plausible assumptions about the values of these hypothetical outcomes.

The "while on treatment" strategy is generally applied to the intercurrent event of treatment discontinuation. This strategy summarises the variable while the patient actually receives their randomised treatment. Results from using this strategy for efficacy data need to be combined with summaries of discontinuations, including the reason for treatment discontinuation and the time to treatment discontinuation. This strategy can also refer to "while at risk" (for example, while alive) if following patients to see if a specific event occurs which may incorporate on and off treatment.

The main difference between the hypothetical strategy and the "while on-treatment" strategy for treatment discontinuation is whether the estimate obtained is considered to apply to the whole scheduled period or only to the period where a patient is taking treatment. For an endpoint measured at specific visits, a hypothetical strategy might seek to estimate the effect at the final visit if all patients had stayed on treatment, a while on-treatment strategy would generally compare a summary across visits while the patient was actually on-treatment.

A principal strata approach to an intercurrent event seeks to restrict the population of interest to the stratum of patients in which an intercurrent event would not have happened (for example, the "principal stratum" of patients who would have completed the treatment of interest). The principal strata strategy is not to be confused with a "complete case" analysis or "per protocol" analysis since these do not restrict both treatment groups in the same way. Hernan and Scharfstein [12] state that "the estimand that compares strategies in a subgroup that cannot be clinically identified … provides a questionable basis for regulatory or clinical decision making". Currently therefore the strategy is not widely used.

Fig. 1 shows how the treatment policy, hypothetical and composite strategies could be applied in practice for a trial in diabetes, with a design similar to the trial for dapagliflozin described above i.e. an endpoint of HbA1c and the key intercurrent events of use of rescue medication. When a treatment strategy is used, rescue medication intake as required could be considered to become an integral part of the treatment regimens under comparison. A composite strategy could involve a definition of a responder as someone with HbA1c $\leq$7% at week 24 without use of rescue medication. For the hypothetical strategy, the assumption could be made that the response without rescue would reflect the response observed immediately prior to rescue.

3.3. Examples of applying different intercurrent event strategies to trials

3.3.1. Applying treatment policy and hypothetical strategies

The PIONEER 1 trial provides an example of use of treatment policy and hypothetical strategies for intercurrent events [13]. This trial compared oral semaglutide to placebo in adult patients with type 2 diabetes. The key intercurrent events were use of rescue medication and discontinuation of randomised treatment [14] and the study used two different estimands.

The first estimand used a treatment policy strategy for both intercurrent events. This estimand assessed the treatment effect regardless of use of rescue medication or discontinuation of trial product and provided a broad perspective of the effect of commencing treatment with semaglutide compared to placebo in the population of patients with type 2 diabetes in clinical practice.

The second estimand used a hypothetical approach for both

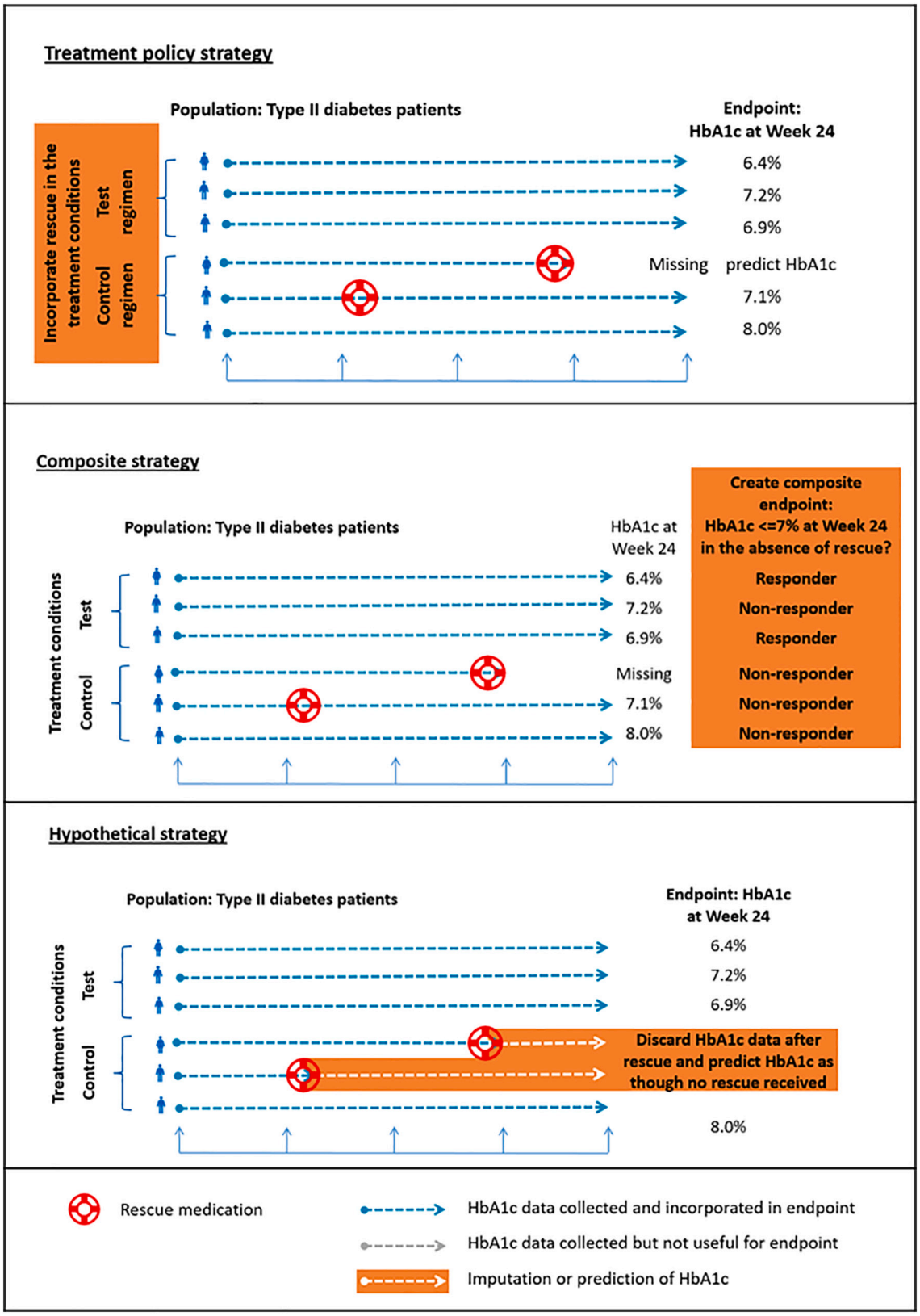

Fig. 1. Illustration of strategies applied to a diabetes trial.

intercurrent events. This estimand assessed the treatment effect if all patients had continued to use trial product for the planned duration of the trial without rescue medication, providing information on the anticipated treatment effect attributable to the medication. Comparative results from the two strategies are provided in Table 4 below.

In the PIONEER example, the two estimands applied the same strategy to the intercurrent events. An example of where different

Table 4
PIONEER 1 study results.

	Placebo ($N = 178$)		Semaglutide ($N = 175$)
Frequency of intercurrent events			
Discontinuation of randomised treatment	19 (11%)		24 (14%)
Use of rescue medication	35 (20%)		7 (4%)
Estimand 1 – treatment policy			
Mean change from baseline HbA1c at week 26	−0.3%		−1.4%
Difference between means (95% CI)		−1.1% (−1.3%, 0.9%)	
Estimand 2 – hypothetical			
Mean change from baseline HbA1c at week 26	−0.1%		−1.5%
Difference between means (95% CI)		−1.4% (−1.7%, 1.2%)	

strategies were used for different intercurrent events is provided by

Degtyarev et al. [2] who describe applying the estimand framework to a study of CAR-T cell therapy in lymphoma. Different estimand strategies were adopted for different intercurrent events (i.e. treatment policy for failure to receive randomised treatment and hypothetical for a new cancer therapy started before the first disease event).

3.3.2. Applying a composite strategy

An example of applying a composite strategy is the SYNAPSE trial which compared mepolizumab to placebo in addition to standard of care for 52 weeks for treatment of chronic rhinosinusitis with nasal polyps [15]. *Co*-primary endpoints were change from baseline in total endoscopic nasal polyp score at week 52 and in mean nasal obstruction VAS score during weeks 49–52.

Occurrence of nasal surgery at any time following randomisation was an anticipated intercurrent event as it was expected to affect evaluation of subsequent scores. A composite strategy was used for this intercurrent event of nasal surgery by incorporating surgery into the definition of these endpoints as a negative outcome.

In this case, there was a clinical rationale for expressing the summary on the original measured scale either by the difference in medians or the difference in means. Use of these summary statistics required a specific arbitrary value for participants with surgery, in this case the worst observed outcome prior to surgery. The key advantage of using medians is that they are largely insensitive to this arbitrary choice of score for patients with surgery while means are dependent on this choice.

The primary estimand was the difference in medians and these were estimated using quantile regression with adjustment for covariates. The approach has been recommended for endpoints where a number of patients are given a "worst score" [11]. A supplemental estimand using a summary of difference in means was completed using MMRM models. Table 5 below shows results for these two estimands.

3.3.3. Applying a principal stratum strategy

There have been some recent published examples of the use of the principal stratum strategy. Qu et al. [16] describe a post-hoc analysis conducted of the IMAGINE-3 study in type 1 diabetes mellitus [17] where this strategy was applied to the intercurrent event of randomised treatment discontinuation.

Magnusson et al. [18] applied the strategy in the context of the EXPAND trial in multiple sclerosis. The primary objective of the trial was to show efficacy of siponimod versus placebo in time to confirmed disability progression, but the question was raised whether a treatment effect would also be present in patients that would not experience relapses. In this setting the relevant intercurrent event is post-randomisation relapse and Magnusson et al. [18] considered estimation of the treatment effect in patients that would not relapse under both randomised treatments.

Table 5
SYNAPSE study results: estimands based on difference in medians and difference in means.

	Placebo ($N = 201$)	Mepolizumab ($N = 206$)
Frequency of intercurrent events		
Surgery for nasal polyps	46 (23%)	18 (9%)
	Primary estimand difference in medians (95% CI)	Supplemental estimand difference in means (95% CI)
Total endoscopic score	−0.73 (−1.11, −0.34)	−0.99 (−1.36, −0.61)
Nasal obstruction VAS score	−3.14 (−4.09, −2.18)	−1.97 (−2.63, −1.31)

3.4. Missing data and sensitivity analysis

It is important to distinguish trial estimands ("what is to be estimated") from trial estimators ("how will the treatment effect be estimated"). Missing data affects trial estimators not trial estimands.

For example, if a treatment policy strategy is employed for use of rescue medication and a patient subsequently decides to withdraw from the study as well as randomised treatment, then data beyond this point is missing data. However, if a hypothetical strategy is employed for use of rescue medication then only data up to this event is used in the analysis; the data after rescue medication, whether collected or not, is not relevant to the estimand. The hypothetical strategy needs to make plausible assumptions about what would have happened during the remainder of the trial had the patient not taken rescue medication, for example that the values would have reflected the patient's previous history as well as values from others who remain in the trial.

It is sometimes claimed that a treatment policy strategy makes fewer unverifiable assumptions than other strategies [9]. However, following study withdrawal, there is no data for the patient after that point. A treatment policy strategy needs to make plausible assumptions about what treatment the patient would receive in the missing observation period and what the outcome would be. The basis for these assumptions may be unclear.

4. Implications for clinical trials

4.1. ITT analysis

Many statisticians have argued that trials should use an ITT approach to analysis, in other words a treatment policy approach to post-treatment events such as rescue therapy use as well as including all randomised patients. For example, Lachin [19] states "the intent-to-treat analysis provides the most realistic and unbiased answer to the more relevant question of clinical effectiveness" and Ellenberg [20] states "the first step in evaluating a therapy must be an unbiased intent-to-treat analysis".

When emphasis is mainly on *p*-values rather than estimation of treatment effects, statistical significance in such an analysis allows reassurance that the treatment under scrutiny has some impact on efficacy under minimal assumptions. However, the estimate of the effect of

a new treatment is important, for example when evaluating risk: benefit and when comparing its cost-effectiveness against alternatives. In this context a conservative estimate may not be appropriate and may not be easily comparable. It should be preferable to answer a question of interest with some assumptions and potential small bias rather than using ITT analysis to estimate a quantity of less interest.

An important concern is whether trial results are generalisable from the trial population to clinical practice. The interpretation of results from an ITT analysis depends on the medications provided when a patient discontinues randomised treatment as well as the randomised medication itself, thus changes in medical practice over time could have a profound effect on the generalizability of results. Ratitch et al. [21] have argued that "treatment policy broadens the treatment regimen under evaluation because it includes whatever treatment is taken (or not taken) … Such loosely defined comparisons are rarely meaningful because the target of inference is not defined and the causal links with the investigated treatment are weakened".

Clinical trials are typically not set up to answer real questions about 'treatment policy', since the study protocol often discourages the use of additional medications or aims to standardize approaches across regions. Akacha et al. have argued that: "If one is really interested in the treatment-policy estimand, then an effectiveness trial, that is, a real world trial design such as a pragmatic study, may be more appropriate" [22].

In the oncology area, crossover of patients assigned to one treatment arm to the other one is a particular problem since "it is often considered unethical not to offer a patient further treatment using an alternative therapy regimen following disease progression" [23]. As Degtyarev et al. state, "the comparison of overall survival between the investigational treatment and the sequence of standard-of-care followed by investigational treatment is, therefore, not assessing a treatment effect of interest" [2]. Since the rescue treatment in this case is investigational product, which is not available to patients, the comparison cannot be considered a valid treatment policy question as neither policy is available as standard of care.

Unfortunately, death is always possible in any trial. Where death is not part of the endpoint, it is important to note that the treatment policy strategy cannot be implemented, since values for the variable after the intercurrent event do not exist.

4.2. Per protocol analysis

Historically analyses have been viewed as a dichotomy: they are either ITT analyses or "per protocol" [24]. The distinguishing feature of Per Protocol analysis is that patients with major protocol violations are excluded altogether and all of their data including data collected prior to the violation is not used. The decision on whether to exclude a patient based on their adherence to the protocol can be somewhat arbitrary and per-protocol analysis can exclude patients who discontinue due to lack of efficacy. As a result of these problems, the role of per-protocol analysis in superiority trials has been relegated and often this analysis is not even performed. A key value of the estimand framework is to provide coherent alternatives to a per-protocol analysis; use of alternatives strategies to treatment policy does not require exclusion of patients who provide poor outcomes on a treatment. For example, the hypothetical strategy could be used to target the treatment effect under the scenario that all patients completed randomised medication to the end of the trial or a principal strata strategy to target the subpopulation of patients who would complete their randomised medication.

4.3. Alternatives: the new estimand framework

A key goal of clinical trials is to enable optimal communication of treatment effects to patients, prescribers and other stakeholders. No patient expects a treatment to work if they do not take it and yet the efficacy results typically presented to them when deciding whether to take a new medication is an average of those who have and those have not taken the medication. Patients and prescribers may be interested in what happens when patients actually take the treatment, which corresponds to a "while on treatment" strategy for treatment discontinuation [24,25]) and/or what happens in patients who take and tolerate the treatment (which corresponds to a principal stratum strategy [22]. Often there is interest in what the treatment effect would be in the absence of specific rescue or alternative medication (which corresponds to a hypothetical strategy).

Darken et al. [26] propose use of an attributable estimand to address intercurrent events. This estimand uses different strategies for different intercurrent events. Those events that are considered to be adversely related to randomised treatment (e.g. discontinuation of treatment due to adverse events or lack of efficacy) are considered attributable and handled with a composite strategy (hence become part of the endpoint of interest), while a hypothetical strategy is used for intercurrent events not considered to be related to randomised treatment (e.g. treatment discontinuation due to administrative reasons such as loss to follow-up).

Use of any of these strategies needs to be part of a tripartite approach [22], including assessment of frequency and reasons for treatment discontinuation due to adverse event and due to lack of efficacy. When these discontinuations favour one treatment compared to another, this needs to be incorporated into the interpretation of the results of the trial.

Discussion of the estimand used for efficacy assessment is relevant for benefit-risk assessment. It is increasingly realised that for events such as treatment discontinuation due to AE, if a treatment policy strategy is used for efficacy and an on-treatment strategy for safety then the event may be "double counted" in the sense that the efficacy is penalised and the event is counted as a safety risk.

4.4. Estimands for COVID-19 risk mitigation

The current pandemic has affected conduct of ongoing clinical trials [21], not only due to direct COVID-19 illnesses but also importantly due to restrictions imposed as a result of the pandemic. For example, patients may have missed doses of randomised treatment in clinical trials. The estimand framework has provided a useful basis for discussions on the impact of COVID-19 on the statistical analysis [27–30].

Most (if not all) ongoing studies were designed to answer a clinical question in the absence of a pandemic. Moving into the future, while the disease itself will still be around, it is likely that much of the impact on daily life will be mitigated. Akacha et al state "A hypothetical estimand strategy seems to be plausible for intercurrent events related to pandemic related operational challenges" [28]. Similarly for estimation in the face of the pandemic it seems appropriate for the statistical analysis to assume that missing data due to the restrictions reflects previous observed values rather than to assume a "worst case" which might be viewed as corresponding to an ITT analysis.

An illustration of the issue is to imagine that one trial is conducted before the pandemic and another is conducted during it. If both studies pre-specified a treatment policy approach for intercurrent events, the pre-defined primary analyses could give quite different answers for the trials due to the second study being impacted by COVID-19. The treatment policy approach implies a specific setting that may not be generalisable in the future. This shows why a hypothetical approach for the second study will be important in evaluating the overall effect of the treatment in a post COVID-19 world.

5. Conclusions

The new estimand framework has allowed discussion of alternatives to the strict ITT approach to statistical analysis. When designing a clinical study, it is important to consider all the key intercurrent events and to decide how these should best be handled for a clinically relevant answer to the scientific question. It is increasingly recognised that the treatment effect estimated by the treatment policy approach may not

always be of primary clinical interest and may not appropriately communicate to prescribers and patients the efficacy that is directly attributable to the treatment. i.e. what can be expected in terms of efficacy if the patients takes the medication as prescribed. An important question for a new medicine is whether it provides an efficacy benefit, another different question is whether treatment strategies result in different long-term outcomes.

Author contribution

ONK is a full-time employee of GlaxoSmithKline and hold shares in the company. DW is a full-time employee of AstraZeneca. AP is a full-time employee of ICON. MW is a full-time employee of Novartis.

Each named author has substantially contributed to drafting this manuscript.

Declaration of Competing Interest

The authors are all employees of pharmaceutical companies or of companies providing services to the pharmaceutical industry and declare that they have no other known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J.A. Lewis, Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline, Stat. Med. 18 (15) (1999 Aug 15) 1903–1942.

[2] E. Degtyarev, Y. Zhang, K. Sen, D. Lebwohl, M. Akacha, L.V. Hampson, B. Bornkamp, A. Maniero, F. Bretz, E. Zuber, Estimands and the patient journey: addressing the right question in oncology clinical trials, JCO Precis. Oncol. 3 (2019) 1–10.

[3] E. Ferrannini, S.J. Ramos, A. Salsali, W. Tang, J.F. List, Dapagliflozin monotherapy in type 2 diabetic patients with inadequate glycemic control by diet and exercise: a randomized, double-blind, placebo-controlled, phase 3 trial, Diabetes Care 33 (10) (2010 Oct 1) 2217–2224.

[4] FDA, Slides for the July 19, 2011 Meeting of the Endocrinologic and Metabolic Drugs Advisory Committee. https://wayback.archive-it.org/7993/20170404152056/https://www.fda.gov/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/ucm264311.htm. Accessed 16 Nov 2020.

[5] B. Holzhauer, M. Akacha, G. Bermann, Choice of estimand and analysis methods in diabetes trials with rescue medication, Pharm. Stat. 14 (6) (2015 Nov) 433–447.

[6] D. Sherifali, K. Nerenberg, E. Pullenayegum, J.E. Cheng, H.C. Gerstein, The effect of oral antidiabetic agents on A1C levels: a systematic review and meta-analysis, Diabetes Care 33 (8) (2010 Aug 1) 1859–1864.

[7] S.D. Aaron, K. Vandemheen, D. Fergusson, et al., Tiotropium in combination with placebo, salmeterol, or fluticasone-salmeterol for treatment of chronic obstructive pulmonary disease: a randomized trial, Ann. Intern. Med. 146 (8) (2007) 545–555.

[8] S.D. Aaron, D. Fergusson, G.B. Marks, et al., Counting, analyzing and reporting exacerbations of COPD in randomized, controlled trials, Thorax 63 (2008) 122–128.

[9] Committee for Human Medicinal Products, ICH E9 (R1) Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials, Final version, Adopted Nov 2019, https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

[10] A. Phillips, T. Clark, Estimands in practice: bridging the gap between study objectives and statistical analysis, Pharm. Stat. 20 (1) (2021 Jan) 68–76.

[11] O.N. Keene, Strategies for composite estimands in confirmatory clinical trials: examples from trials in nasal polyps and steroid reduction, Pharm. Stat. 18 (1) (2019 Jan) 78–84.

[12] M.A. Hernan, D. Scharfstein, Cautions as regulators move to end exclusive reliance on intention to treat, Ann. Intern. Med. 168 (2018) 515–516.

[13] V.R. Aroda, J. Rosenstock, Y. Terauchi, Y. Altuntas, N.M. Lalic, E.C. Villegas, O. K. Jeppesen, E. Christiansen, C.L. Hertz, M. Haluzík, PIONEER 1: randomized clinical trial of the efficacy and safety of oral semaglutide monotherapy in comparison with placebo in patients with type 2 diabetes, Diabetes Care 42 (9) (2019 Sep 1) 1724–1732.

[14] V.R. Aroda, T. Saugstrup, J.B. Buse, M. Donsmark, J. Zacho, M.J. Davies, Incorporating and interpreting regulatory guidance on estimands in diabetes clinical trials: the PIONEER 1 randomized clinical trial as an example, Diabetes Obes. Metab. 21 (10) (2019 Oct) 2203–2210.

[15] J.K. Han, C. Bachert, W. Fokkens, M. Desrosiers, M. Wagenmann, S.E. Lee, S. G. Smith, N. Martin, B. Mayer, S.W. Yancey, A.R. Sousa, Mepolizumab for chronic rhinosinusitis with nasal polyps (SYNAPSE): a randomised, double-blind, placebo-controlled, phase 3 trial, Lancet Respir. Med. (2021 Apr 16), https://doi.org/10.1016/S2213-2600(21)00097-7.

[16] Y. Qu, J. Luo, S.J. Ruberg, Implementation of tripartite estimands using adherence causal estimators under the causal inference framework, Pharm. Stat. 20 (1) (2021 Jan) 55–67.

[17] R.M. Bergenstal, H. Lunt, E. Franek, F. Travert, J. Mou, Y. Qu, C.J. Antalis, M. L. Hartman, M. Rosilio, S.J. Jacober, E.J. Bastyr III, Randomized, double-blind clinical trial comparing basal insulin peglispro and insulin glargine, in combination with prandial insulin lispro, in patients with type 1 diabetes: IMAGINE 3, Diabetes Obes. Metab. 18 (11) (2016 Nov) 1081–1088.

[18] B.P. Magnusson, H. Schmidli, N. Rouyrre, D.O. Scharfstein, Bayesian inference for a principal stratum estimand to assess the treatment effect in a subgroup characterized by postrandomization event occurrence, Stat. Med. 38 (23) (2019 Oct 15) 4761–4771.

[19] J.M. Lachin, Statistical considerations in the intent-to-treat principle, Control. Clin. Trials 21 (2000) 167–189.

[20] J.H. Ellenberg, Intent-to-treat analysis versus as-treated analysis, Drug Inform. J. 30 (1996) 535–544.

[21] B. Ratitch, J. Bell, C. Mallinckrodt, J.W. Bartlett, N. Goel, G. Molenberghs, M. O'Kelly, P. Singh, I. Lipkovich, Choosing estimands in clinical trials: putting the ICH E9 (R1) into practice, Therap. Innov. Regul. Sci. 54 (2) (2020 Mar) 324–341.

[22] M. Akacha, F. Bretz, S. Ruberg, Estimands in clinical trials – Broadening the perspective, Stat. Med. 36 (2017) 5–19, https://doi.org/10.1002/sim.7033.

[23] P. Tappenden, J. Chilcott, S. Ward, S. Eggington, D. Hind, S. Hummel, Methodological issues in the economic analysis of cancer treatments, Eur. J. Cancer 17 (2006) 2867–2875.

[24] O.N. Keene, Intent-to-treat analysis in the presence of off-treatment or missing data, Pharm. Stat. 10 (2011) 191–195.

[25] O.N. Keene, S. Ruberg, A. Schacht, M. Akacha, R. Lawrance, A. Berglind, D. Wright, What matters most? Different stakeholder perspectives on estimands for an invented case study in COPD, Pharm. Stat. 19 (2020) 370–387.

[26] P. Darken, J. Nyberg, S. Ballal, D. Wright, The attributable estimand: a new approach to account for intercurrent events, Pharm. Stat. 19 (5) (2020 Sep) 626–635.

[27] R.D. Meyer, B. Ratitch, M. Wolbers, O. Marchenko, H. Quan, D. Li, C. Fletcher, X. Li, D. Wright, Y. Shentu, S. Englert, Statistical issues and recommendations for clinical trials conducted during the COVID-19 pandemic, Stat. Biopharm. Res. 6 (2020) 1–22.

[28] M. Akacha, J. Branson, F. Bretz, B. Dharan, P. Gallo, I. Gathmann, R. Hemmings, J. Jones, D. Xi, E. Zuber, Challenges in assessing the impact of the COVID-19 pandemic on the integrity and interpretability of clinical trials, Stat. Biopharm. Res. 17 (2020 Aug) 1–8.

[29] B.C. Kahan, T.P. Morris, I.R. White, C.D. Tweed, S. Cro, D. Dahly, T.M. Pham, H. Esmail, A. Babiker, J.R. Carpenter, Treatment estimands in clinical trials of patients hospitalised for COVID-19: ensuring trials ask the right questions, BMC Med. 18 (1) (2020 Dec) 1–8.

[30] C.U. Kunz, S. Jörgens, F. Bretz, N. Stallard, K. Van Lancker, D. Xi, S. Zohar, C. Gerlinger, T. Friede, Clinical trials impacted by the COVID-19 pandemic: adaptive designs to the rescue? Stat. Biopharm. Res. 12 (4) (2020 Oct 1) 461–477.