

# Intention-to-Treat Analysis but for Treatment Intention: How should Consumer Product Randomized Controlled Trials be Analyzed?

Rolf Weitkunat\*, Gizelle Baker and Frank Lüdicke

*Philip Morris Products SA, Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland*

**Abstract:** *Background:* Experimental study design, randomization, blinding, control, and the analysis of such data according to the intention-to-treat (ITT) principle are de-facto “gold standards” in pharmacotherapy research. While external treatment allocation under conditions of medical practice is conceptually reflected by in-study randomization in randomized controlled trials (RCTs) of therapeutic drugs, actual product use is based on self-selection in a consumer product setting.

*Discussion:* With in-market product allocation being consumer-internal, there is no standard against which protocol adherence can be attuned, and the question arises, as to whether compliance-based analysis concepts reflect the real-world effects of consumer products.

*Summary:* The lack of correspondence between RCTs and consumer market conditions becomes evident by the fact that even if, theoretically, all data would be available from all members of the real-world target population, it would be impossible to calculate either an ITT or a per-protocol effect. This renders the calculation of such estimates meaningless in consumer product research contexts.

**Keywords:** Randomization, self-selection, intention-to-treat, actual use, consumer products.

## BACKGROUND

The effects of drug-based medical interventions are generally assessed in randomized controlled trials (RCTs), by performing statistical comparisons of the outcome measures, between patients randomly allocated to the investigational treatment ( $R=1$ ) or to the control treatment ( $R=0$ ). Under the assumptions of (1) balance of both the known and unknown covariates (through randomization); (2) full compliance; and (3) complete data, the results of the trial are evidence of a causal treatment effect. This is a fundamental strength of the randomized-experimental methodology, in terms of internal validity [1]. In potential outcomes terms, randomization renders the impact of the actual treatment ( $A$ ) on the distribution of the potential outcomes  $Y(A=a)$  ignorable (i.e.,  $Y(a) \perp A$ ) and participants exchangeable across groups [2,3].

In a well-designed RCT, the factual probabilistic dependencies of an outcome measurement on the actual treatment can be interpreted in a causal manner in terms of  $A$  having caused  $Y$  in at least some individuals. This probability-raising formulation of causation, which is restricted to the study-specific context ( $C$ ), has been formulated for observed 0/1 response outcomes through the conditional probability relation  $\Pr(Y|A=1,C) > \Pr(Y|A=0,C)$  [4]. When this inequality is demonstrated, the probabilistic causal

conclusion that  $A$  caused  $Y$  is implicated within this study context. The external validity of the conclusions is dependent on the context in which the study was performed, therefore the findings are only generalizable when the study context reflects the context in which the treatment will be used. In reality this can only be approximated. Importantly, as Cartwright [5] has pointed out, there are “no manuals” for the type of induction implied in generalizing the results from a specific study. This is mainly due to the lack of certainty regarding the totality of relevant context factors.

Even if researchers were able to design and conduct an “ideal” RCT, clinical trials become more complex in reality because they involve human participants who deviate from the study instructions [6]. In fact, compliance is a complicated and typically non-random phenomenon which encompasses different forms of “all-or-nothing” and partial compliance. The type and extent of non-compliance that arise in a study depends on the research design and study context, and can range from not taking the allocated treatment, to not taking the treatment in accordance with the protocol, or even taking an alternative treatment [7]. The intuitive approach to addressing the issues that arise with a broken randomization, when the actual treatment  $A_i$  assigned to the  $i^{\text{th}}$  participant is not the randomized treatment  $R_i$ , would be to simply restrict the analysis to the individuals that completed the study as planned. This equates to a “per-protocol” (PP) analysis, which does inform about the efficacy of the experimental treatment compared to the control under ideal circumstances. It can provide “proof” of a

\*Address correspondence to this author at the Philip Morris Products SA, Quai Jeanrenaud 5, 2000 Neuchâtel, Switzerland; Tel: +41 58 242 2563; Fax: +41 58 242 2811; E-mail: rolf.weitkunat@pmi.com

therapeutic concept by answering the question “can the treatment work” [8], for a specific study context, by demonstrating  $E[Y|A=1, R=1] > E[Y|A=0, R=0]$ .

### Effect Estimation in RCTs

Unfortunately, PP effect estimates can be biased, because the groups being compared are not only based on randomized treatment allocation, but also on post-randomization compliance. If compliance is not ignorable [9], then the factors that determine treatment compliance can also influence the effect of the treatment and thus confound the treatment-effect association. Also, if compliance is influenced by both the treatment and the treatment effect, the PP approach might overestimate the benefit of the treatment, because the target population will also include less responsive individuals. Therefore, while the PP analysis may have the advantage of alleviating the need to analyze and address the details of non-compliance, the dismissal of incomplete and non-compliant data implies a loss of information and power, and can potentially introduce bias by violating the ignorability assumption.

In non-randomized, observational cohort studies the “as-treated” (AT) analysis, where subjects are analyzed according to treatment they received, is the only viable analysis. This approach is also used when assessing safety data collected in RCTs and is the standard approach for analyzing preventive vaccine trials [10]. In order to correctly specify actual treatment (i.e., the exposure) for an AT analysis, the details of non-compliance need to be analyzed, to determine whether or not the treatment was taken in accordance with the protocol (e.g., dose, frequency), or whether it was declined, replaced, or supplemented with an alternative treatment.

When compliance is independent of baseline characteristics and treatment effects, the PP and AT analyses provide valid effect estimates, as the assumption of ignorability holds. However, when the randomization cannot be relied on, ignorability is no longer warranted and the statistical analysis needs to consider the impacts of potential biases and confounding. In this scenario non-compliance is likely linked to a diminished treatment effect and/or an adverse outcome, which means that the AT and PP analyses would be prone to overestimate the treatment effect. Methods exist that can be used to control for confounding in effect estimates analytically, by reconstituting conditional exchangeability ( $Y(a) \perp A|C$ ),

or, in terms of observed outcomes, conditional independence through unconfounding, such that  $Y \perp R|(A,C)$  [11,12]. This can be achieved by implementing some form of conditioning (adjustment, stratification, standardization, or matching) to account for potential confounders. However, there is no guarantee that unconfounding is being achieved, and unmeasured or unknown confounders cannot be considered.

To avoid the methodological complications associated with the PP and AT analyses, which arise because of the dissociation of randomized from actual treatment, the intent-to-treat (ITT) analysis principle can be followed. This is where the analysis is performed according to the randomization, independent of the post-randomization compliance and actual treatment. The approach alleviates the need to exclude certain subjects from the analysis or to re-define the comparison groups, by simply demonstrating that for a specific study context  $E[Y|R=1] > E[Y|R=0]$ . While the ITT analysis may appear counter-intuitive at first, it has the advantage of building on the randomization-based balance of the baseline covariates, as well as retaining the information from the full study sample.

When noncompliance arises because of a lack of treatment effect, the ITT estimates are conservative and minimize the type I error. As non-compliance is not only a study phenomenon but also occurs in medical practice, ITT results are widely considered to be more generalizable to the real-world than PP or AT results. Additionally, the ITT analysis has the potential to minimize investigator bias and thereby increase the integrity of study results. As pointed out by Begg [13] the ITT analysis promotes transparency by severely limiting opportunities to obscure unfavorable results. It is understandable that these methodological strengths, together with the statistical simplicity, have led to widespread adoption of the ITT approach, making it the “gold standard” approach in analyzing randomized superiority trials and have led to broad support by regulatory bodies [7,14]. Table 1 provides an overview of ITT as compared to PP and AT analysis strategies in the context of randomized parallel-group controlled therapeutic superiority trials.

### Scope and Properties of ITT

Although there are many advantages to the ITT approach, there are also downsides. A treatment which is efficacious may not be effective (because of non-compliance), which is why exclusive use of the ITT

**Table 1: Key Aspects, Strengths and Limitations of Main Analytical Approaches in Superiority Therapy Studies. (Characteristics do not Simply Extend to Non-Inferiority Assessment)**

	Analytical Approach		
	Intention to Treat	Per Protocol	As Treated
Basic principle of comparing participants	As randomized, ignoring actual treatment	As randomized, conditional to protocol compliance	As per treatment actually received, ignoring randomization
Scope	Effectiveness of treatment offer	Efficacy of treatment under ideal circumstances (compliance)	Efficacy of treatment; safety
Properties	Typically underestimating superiority effect	Ideal effect estimated in superiority assessment (anti-conservative in superiority assessment)	Estimated effect subject to self-selection bias, typically anti-conservative in superiority assessment; unbiased effect estimation may require conditioning (e.g., adjustment)
Strengths	Randomization-protected from bias due to imbalance of baseline characteristics; simple	Proof of therapeutic concept	Data set comparable to that of safety analysis and to observational studies; allows for analysis under high treatment crossover rates
Limitations	Imputation of missing data required; generalizability depending on correspondence of in-study and real-life compliance	Reduced power depending on non-compliance; risk of ignorability assumption being violated (selection bias); bias possible in any direction, most likely anti-conservative	Reduced power depending on dropout-rate; bias possible in any direction, most likely anti-conservative

approach does not provide sufficient insight into the actual effect of a treatment. While confounding due to self-selection subsequent to randomization is avoided as compliance is ignored, ITT effect estimates of parallel-group superiority trials are not independent of compliance, as they are biased towards the null with increasing rates of noncompliance. Moreover, the ITT strategy is inappropriate for safety analyses [15].

Because post-randomization drop-out may be linked to a lack of benefits or tolerability of the treatment, adjustment to avoid bias due to differential loss to follow-up is needed for conducting an ITT analysis. One of the simplest adjustments is the last observation carried forward (LOCF) approach, a single imputation method where the missing data are replaced by the last observed value. Although this approach was previously considered to be conservative, it introduces bias and leads to overestimation of the precision of the effect estimate [16]. While there are statistical methodologies that can reduce the potential bias that arises from missing data (e.g., multiple imputation, generalized estimating equations methods) [17], ITT estimates remain at risk of being seriously biased when more than 30 percent of study participants are lost to follow-up [18].

Although ITT estimates are usually smaller than PP and AT estimates in superiority trials, the conservativeness does not extend to non-inferiority or

equivalence trials, where the ITT approach tends to favor equality of treatments and therefore to increase the type I error. Furthermore, the ITT approach, by mixing the effect in compliers with the absence of effect in non-compliers, dilutes the treatment effect estimates and increases the type II error in superiority trials.

In addition to their general strengths and limitations, one has to consider the research questions that are addressed by the ITT and non-ITT approaches. The PP analysis answers questions about the maximal treatment effect or the effect when the treatment is received as intended, whereas the AT analysis answers a different question, related to the effect of the treatment, as it is actually received. Both provide measures of efficacy aiming at explaining the treatment effects. In contrast, the ITT analysis aims to quantify the effect of being "prescribed" a treatment, regardless of whether or not the treatment is received. Therefore, ITT estimates pragmatically quantify the effectiveness of being prescribed a treatment, and do not directly address the efficacy of the treatment. Assessing the effects of treatment assignment rather than treatment *per se* has in fact been considered an asset with regard to similarity to the real-world medical practice and its value for informing policy decisions. However, generalizability of ITT results relies on the assumption that the levels and patterns of noncompliance in the study and real-world conditions are similar.

Consequently, the properties of ITT estimates need to be carefully assessed in context. This is particularly important when comparing results from RCTs to observational studies [19]. As Hernán and Robins [20] have pointed out, reporting only the ITT effect implies preferring misclassification over confounding, a preference that needs to be justified in each application.

The ITT principle was developed for and is inextricably linked to research on medical treatments. And in this framework, the approach is reasonable as the treatment allocation is externally controlled and can be considered to mirror the real-world medical practice of prescribing treatments to patients. In therapy research this link is rarely addressed, due to the similarities between the study and real-world setting, and seems to be generally viewed as a triviality. The concept, as recognized by the term “intention” in the name ITT, reflects an external intent to impose a treatment on a subject, which in RCTs is operationalized by randomization.

Even if ITT estimates reflect the real-world effectiveness of interventions, it is important not to confuse them with estimates of the population health impact. ITT quantifies the effectiveness at an individual-level, whereas population impact measures weigh exposure effects by the proportion of the total population that is exposed to the intervention. To obtain valid population impact estimates, calculations must be conducted in accordance with the exposure and covariate strata that actually define the target population. If, for example, the effect of a therapy on an outcome depends on sex, age, dose, or other factors, then stratum-specific effect estimates as well as stratum-specific actual treatment prevalence estimates must be obtained in order to quantify the population attributable impact [21].

**DISCUSSION**

**Transportability of Effect Measures**

In order to fully understand the effect of a treatment one would have to be able to assess both of the two potential outcomes for a subject (1) the potential outcome of receiving the treatment  $Y(1)$  and (2) the potential outcome of not receiving the treatment  $Y(0)$ , where the difference in the two potential outcomes constitutes the individual treatment effect  $ITE_i = Y_i(1) - Y_i(0)$ . Assuming full compliance ( $A_i = R_i$  for all participants  $i$ ), randomization can be thought of as a

process that, by creating one missing potential outcome for each participant, generates the observed outcomes of the analysis variable so that  $Y_i = A_i \cdot Y_i(1) + (1 - A_i) \cdot Y_i(0)$  [22]. While these do not allow for calculating the counterfactual individual or average treatment effect  $ATE = E(ITE)$ , the average observed treatment effect can be quantified by a contrast of the expected values of  $Y$  under the two exposures,  $\Delta Y = E[Y|A=1] - E[Y|A=0]$ , which is unbiased under full compliance. When the distribution of the observed  $Y$  is normal or Bernoulli,  $\Delta Y$  can be calculated as the mean difference,  $\Delta Y = M_1 - M_0$ . Under the assumption of a sharp null effect in  $Y_1$  with no actual treatment ( $A=0$ ; meaning  $M_0=0$ ), this simplifies to

$$\Delta Y = M_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} Y_i | (A = 1) \tag{1}$$

$\Delta Y$  corresponds to the average observed treatment effect in the treated, (i.e., the as-treated estimate). When the proportion of the total sample that is actually receiving treatment ( $P_A$ ) corresponds to

$$P_A = \Pr(A = 1) = \frac{\sum_{i=1}^N (A_i = 1)}{\sum_{i=1}^N (R_i = r)} \tag{2}$$

it follows that  $N_1 = P_A \cdot N$  in (1).

By building directly on  $P_A$ , the AT estimate can be calculated without conditioning the summation, as in (1), through

$$AT = \frac{\frac{1}{N} \sum_{i=1}^N Y_i}{P_A} \tag{3}$$

Then based on the proportion of the sample that is randomized to active treatment ( $P_R$ )

$$P_R = \Pr(R = 1) = \frac{\sum_{i=1}^N (R_i = 1)}{\sum_{i=1}^N (R_i = r)} \tag{4}$$

and on the control group not having access to the treatment (one-sided non-compliance), the ITT effect is calculated as

$$ITT = \frac{\frac{1}{N} \sum_{i=1}^N Y_i}{P_R} \tag{5}$$

while the PP effect is similarly calculated as

$$PP = \frac{1}{N} \sum_{i=1}^N Y_i \quad (6)$$

$$P_R \cdot P_C$$

using the protocol-compliant proportion ( $P_C$ ) of the sample

$$P_C = E(A = z | R = z) = \Pr(A = 1 | R = 1) = \frac{\sum_{i=1}^N (R_i = 1, A_i = 1)}{\sum_{i=1}^N (R_i = 1)} \quad (7)$$

which corresponds to the ratio of the unconditional probabilities  $P_A$  and  $P_R$  under one-sided non-compliance.

The equations for the three estimates: the AT estimate (3), the ITT estimate (5), and the PP estimate (6) are computationally identical in the numerator, while the factors related to randomization, compliance, and actual treatment are included in the denominators. Although it may appear that the numerators are independent of randomization and compliance, they are not. The expectancy is the counterfactual average treatment effect in the treated  $ATT = E(ATE | A = 1)$ , weighted by the proportion of the sample that is actually treated, i.e.,  $E[N^{-1} \cdot \sum Y_i] = ATT \cdot P_A$ , with  $P_A = P_R \cdot P_C$ . Under one-sided non-compliance, the complier average causal effect (CACE) equals the average treatment effect in the treated (ATT).

By replacing the numerator in (5) accordingly, it follows that  $ITT = CACE \cdot P_C$ , which implies that the ITT estimate is dependent on both the CACE and the level of compliance, even though, as pointed out above, it should not be biased by self-selection. In contrast to ITT, PP and AT are treatment effect estimates which are not biased by compliance level per se, even though they can be confounded by differential non-compliance or common causes (or effects) of actual exposure and effect.

The dependence of ITT estimates on compliance levels can also be demonstrated without the use of counterfactual concepts. Constraining  $Y$  to follow a Bernoulli distribution with  $E(Y | A = 1) = 1$ , assuming one-sided non-compliance and a sharp null effect with no actual treatment, the expectancy of the sum  $E[\sum Y_i] = N \cdot E[Y] = N \cdot P_A = N \cdot P_R \cdot P_C$ , implying that the numerator can be simplified to  $P_A = P_R \cdot P_C$ . Further, it follows that  $AT = PP = 1$ ,  $ITT = P_C$ , and that  $P_R$  has no impact on the estimates. Thus, under full compliance, all three estimates are identical. This simplifies the assessment of the external validity [1], i.e., the generalizability of study findings to the real therapy

world. If minor sources of divergence between in-study vs. medical practice realities are ignorable, four major assumptions are needed to render effect estimates unbiased: (1) the study population is an unbiased representative sample of the target patient population; (2) the healthcare providers do not introduce treatment selection bias; (3) the compliance levels are identical under study and medical practice conditions, (4) the study and real-world conditions do not differentially influence the treatment effects. In reality, under most circumstances, the first three assumptions are more fiction than fact.

### Consumer Product Effects

As concluded above, the basis for designing a study relevant to the real-world is that a selection mechanism can be implemented to obtain a study sample that is representative of the target population. In drug trials the mechanism of recruiting and screening subjects is remarkably similar to the real-world medical practice of diagnosing and treating an ailment. Both settings rest on confirmed medical diagnosis and restrictions regarding co-morbidities, concomitant medication and access to the investigational treatment itself. All of this is in favor of a relatively close correspondence between study and real-world conditions. Also, patients typically are in need of and expect a benefit from a treatment to which they have restricted access, whereas consumers rarely depend on any particular product and typically can access alternative products freely. This renders participation and compliance rates likely higher in therapy as compared to consumer product trials. Thus, even if recruitment for and conduct of studies are similar, drug trials are likely to be more representative of their target populations and settings than consumer product trials.

When it comes to transporting findings from consumer product RCTs to consumer markets, things become very complicated. Blinding or concealing the actual exposure is often difficult or impossible for consumer products and may, if achieved, impact the compliance, as the subjects would be unable to identify with the product as they do in real-world settings. For studies where the control product is already marketed, and the study is being conducted in an ambulatory setting, there are further complications related to non-compliance. Study participants often don't have access to the investigational product (as it is not yet marketed), but they have free access to the control product or even to other alternative products. And although randomization is used as a mechanism to allocate product, this mechanism does not reflect the self-

selection process of choosing a new consumer product in the market. As pointed out above, ITT effect estimates can only be generalizable if one assumes that compliance in a setting of random consumer product allocation is reflective of the consumer product selection in the market. And even then it would need to be acknowledged that the usual conservative property of ITT would only hold if in-study compliance would not be larger than real-world actual use rates. As this appears to be unlikely for most settings and products, the ITT approach would overestimate the real-world effectiveness of consumer products.

Further, since with consumer products there is no external “treatment” allocation, both the PP and ITT have no correspondence with the real-world. This raises, in the context of consumer product clinical trials, a fundamental problem, i.e., the necessity for an in-study effect estimate to represent something meaningful in a real-world consumer product market.

It appears that the problem of incorporating external allocation in the consumer product setting has been largely neglected thus far in many consumer-related research areas, although not completely. Hoxby and Murarka [23] point out that the effects of school admission offers are not adequately being dealt with in terms of ITT effect estimates, as final enrolment was always voluntary and non-enrolment would not correspond to noncompliance in therapy settings. According to Feinman [24], ITT analyses in nutrition studies are frequently conducted without justification. These examples seem to demonstrate violations of a basic principle of biostatistics, namely to invariably align the analytic approach with the research question, or, to cite Sackett and Wennberg [25]: “*the question being asked determines the appropriate research architecture, strategy, and tactics to be used – not tradition, authority, experts, paradigms, or schools of thought*” (p. 1636).

As PP and ITT are not applicable for measuring the effects of actual use of consumer products, the question about the meaning of AT is of key importance in consumer studies. The target population-level actual use effect (AU) can, in theory (with  $N_{TP}$  reflecting the size of the target population), be quantified along the same lines as the AT effect under study conditions:

$$AU = \frac{\sum_{i=1}^{N_{TP}} Y_i}{\sum_{i=1}^{N_{TP}} (A_i = 1)} = \frac{\frac{1}{N_{TP}} \sum_{i=1}^{N_{TP}} Y_i}{\frac{1}{N_{TP}} \sum_{i=1}^{N_{TP}} (A_i = 1)} = \frac{\frac{1}{N_{TP}} \sum_{i=1}^{N_{TP}} Y_i}{P_A} \quad (8)$$

As can be seen, the effect estimated in (8) formally corresponds to (3), which estimates AT from study data. If the numerator of (8) is replaced by its expectancy ( $CACE \cdot P_A$ ) it becomes evident that AU is the real-world observable pendant to the average treatment effect in the treated, ATT. Other than the in-study AT effect estimate, AU is by definition unbiased, as it reflects the post-launch reality using, theoretically, the totality of data on the target population, rendering it a measure rather than an estimate. A closer look at the denominator of (8) reveals that, through simple counting, the level of actual use is quantified as an unconditional prevalence. From formulas (3), (5), and (6), it becomes clear that the only effect estimate from an RCT that can be obtained based only on simply counting actual use is AT. In contrast, counting conditioned on randomization is required for PP and ITT, respectively. While (8) can be helpful when separate data sets on outcome and actual use rates are available for a particular population, typically population-level AU cannot realistically be measured by a single study; more practically, population-based observational studies can attempt to estimate AU effects by measuring outcomes in representative actual user cohorts. The calculation of the AU effect would then correspond to the standard way of calculating cohort effects, based on only the numerator of (8) and  $N_{TP}$ , the latter then corresponding to the size of the actual use cohort.

While formally, AT and AU are identical, it is critical to note that the causal factors that determine actual use are different under RCT and consumer market conditions. Actual use is based on unconditional counting in the consumer product market situation or in a representative cohort of actual users, but is defined under randomized product use allocation by  $P_A = P_R \cdot P_C$  in an RCT with one-sided non-compliance. In order to render RCT-based AT a valid estimate of AU, it would be required that selection mechanisms, participant/consumer characteristics, compliance/continuation of product use, actual exposure/actual use and effects are identical under study and market conditions. This essentially necessitates that what would be achieved by randomization is an unbiased version of self-selection under market conditions, which inevitably raises the question as to whether randomization (and, more generally, external product use allocation) is a useful concept in consumer product clinical trials in the first place. It appears that this question needs to be addressed very carefully on a case-by-case basis. The longer the trial and the lesser the control over actual

use in the context of a particular research question, the weaker the argument in favor of randomization becomes.

## SUMMARY

ITT is a methodological development in pharmacotherapy research, where it is now widely adopted, and any departure is viewed with suspicion by “clinical trialists” [26] who are in favor of and accustomed to the method. According to Sackett [27], the main purpose of RCTs regarding “the clinical management of patients” is to avoid confusing efficacy with compliance, regression to the mean, or placebo effects. The keystone is randomization, the essential idea of which is to ensure that participants differ only with respect to the interventions being compared. The underlying logic is most likely to play out when RCTs are short, small, double-blinded, well-controlled, and conducted in homogeneous, highly selected and fully compliant patients. Whenever research questions, study designs, and procedures deviate from the textbook-ideal pharmacotherapy RCT, however, the underlying logic deteriorates. When the ITT approach is transplanted to a fundamentally different setting, matters get even worse.

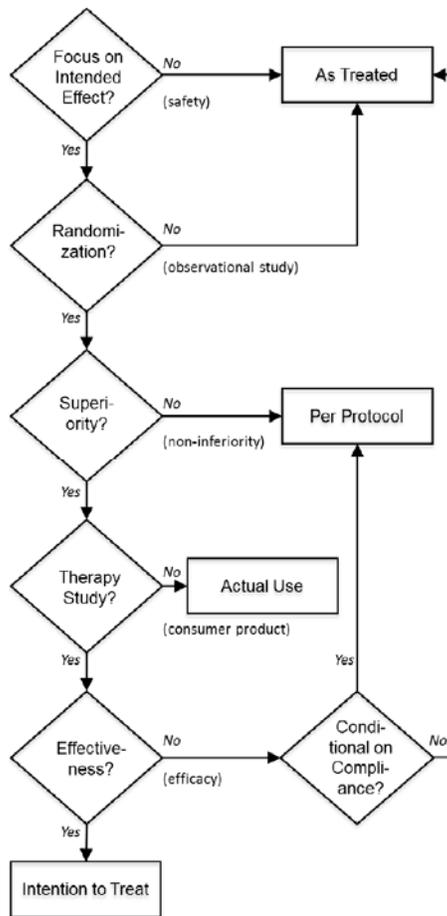
Randomization has no correspondence in real-world consumer product settings with unrestricted product access and self-selection rather than external allocation, and blinding often being impossible. Under consumer product market conditions there is no restricted access to products, no patient, no doctor, no diagnosis, no disease, no treatment, no treatment intention, no treatment allocation, no compliance, and also no noncompliance. Since no prescriptive product allocation takes place in consumer markets, a post-launch equivalent of PP, ITT, or CACE (all of which are inseparable of randomization or at least external treatment allocation) is undefined and cannot be calculated even when *all* target population data are available. As Shrier *et al.* [28] have pointed out recently, “...adherence concepts do not apply when participants simply choose treatment because there is nothing to adhere to” (p.2). This implies that compliance-based and ITT effect estimates are out of context and meaningless with regard to estimating consumer market actual use effects. An inevitable logical consequence is to abandon the ITT principle in consumer product trials.

In contrast, a population health impact assessment based on “as-used” (“as-treated”) effect estimates

appears to be informative. AU estimates should be derived from studies where exposure-response data have been obtained for all strata of relevance, in particular with regard to various levels of dose, as they occur in the real-world, along with prevalence data regarding the size of all strata of relevance in the total population. The heterogeneity of actual use effects needs to be carefully investigated by providing *per actual-use* estimates for subgroups of actual use as well as user characteristics. Based on the probability-raising framework for binary variables, these subgroup-specific effect estimates conditional on sets of subgroup-defining factors (contexts C) can be expressed as  $[\Pr(Y|A)-\Pr(Y|\neg A)]/C$ . By combining these *per actual-use* effect estimates per stratum (AUS) with the population prevalence rates of the same strata, population health impact can be quantified. The so derived public health impact estimation would be grounded solely in study-based efficacy measures as well as in actual-use as it indeed occurs in the target population. Figure 1 provides a generic flowchart on analysis strategies for effect estimates from parallel group therapy as well as consumer product clinical studies.

It appears noteworthy that for purposes of population health impact assessment it is not required that an in-study AUS effect estimate allows for a causal, i.e., unbiased efficacy-like interpretation within the population stratum it corresponds to. For example, if product use is accompanied by favorable lifestyle changes (concomitant effective exposures), the AUS estimate can still inform about the real-world total effect, necessarily under the condition that the concomitant in-study changes correspond to what occurs in the population under natural conditions; whenever synergistic lifestyle changes are more likely to occur in the natural setting compared to a study setting, the AUS estimate will in fact even be conservative with respect to the real-life impact of the product. To the degree that all relevant concomitant lifestyle patterns and other confounders can and in fact are measured (observed observables), these assumptions can be empirically validated through post-market surveillance studies. Avoiding the need to identify the direct (attributable) causal effect of actual product use on the outcome in the context of a public health-level interpretation implies that AUS, so defined, factually provides an effectiveness estimate.

In spite of the fundamental importance of the above methodological questions, they have so far attracted only limited attention in the consumer research realm.



**Figure 1:** Main analysis strategies of parallel-group clinical studies. For unbiased effect estimation, AT and PP estimates require some form of conditioning if confounding cannot be excluded. ITT analyses typically require missing data imputation and provide pragmatic estimates of treatment allocation effects, not of treatment (efficacy) itself. Generalization of ITT results rests on the close correspondence of in-study and real-world patterns and levels of non-compliance. The flowchart represents a generic approach and does not replace context-specific analysis planning.

Not infrequently unscholarly enforcement of the ITT approach, often by non-experts, has reached a concerning level of complacent imprudence, and in order to comply with editorial, granting, or regulatory requirements, researchers are essentially forced to perform such an analysis. Under such circumstances, one viable route could be to lay out clearly what question is and is not actually addressed with this type of analysis in a given context, and to conduct, if possible, additional analyses that provide better or supplementary insights. Even authors that are generally in favor of ITT as the primary approach in clinical therapy research have pointed out the importance of being able to provide advice to patients based on estimates of treatment efficacy under compliance [29]. Furthermore, in the context of clinical therapy trials, sensitivity analyses have been proposed

to assess the effects of non-compliance with the allocated treatment as well as of other forms of protocol deviations; typically, ITT-based treatment effect estimates are compared to PP and/or AT estimates [30]. Along the same lines, the robustness of findings from clinical consumer product studies can be addressed by such sensitivity analyses.

ITT has been developed and instituted by epidemiologists and biostatisticians to address the demand of being able to analyze therapy experiments in humans, i.e., double-blind randomized clinical pharmacotherapy trials with noncompliance and missing data, in an unbiased manner. Whenever applied with care within this framework, ITT undoubtedly is and remains a methodological cornerstone of evidence-based medicine. However, the further research questions and contexts deviate from the original framework, the less self-evident its deployment becomes. When blinding or masking of treatments is impossible and randomization does not reflect real-life consumer choices, careful consideration needs to be given as to what methodologies to apply, and at what evidentiary level. In any case, familiarity with a reiterated concept should not be confused with its universal truth and applicability.

**DECLARATION OF INTERESTS**

RW, GB, and FL work for Philip Morris International (PMI), R&D, receive their salary from PMI and own PMI shares.

**AUTHOR’S CONTRIBUTIONS**

The manuscript evolved from discussions among all three authors. RW drafted the manuscript and GB and FL reviewed multiple versions and provided major contributions to a series of revisions. All authors read and approved the final manuscript.

**ACKNOWLEDGEMENTS**

All costs of this project were covered by Philip Morris International. The views expressed are those of the authors.

**REFERENCES**

[1] Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally 1963.  
 [2] Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized treatments. J Educ Psychol 1974; 66: 688-701.  
<http://dx.doi.org/10.1037/h0037350>

- [3] Rosenbaum P, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70: 41-55.  
<http://dx.doi.org/10.1093/biomet/70.1.41>
- [4] Fitelson B, Hitchcock C. Probabilistic measures of causal strength. In: Illari PM, Russo, R, Williamson J, editors. *Causality in the sciences*. Oxford: Oxford University Press 2011; pp. 600-27.  
[http://dx.doi.org/10.1093/acprof:oso/9780199574131.003.002\\_9](http://dx.doi.org/10.1093/acprof:oso/9780199574131.003.002_9)
- [5] Cartwright N. Knowing what we are talking about: why evidence doesn't always travel. *Evid Policy* 2013; 9: 97-112.  
<http://dx.doi.org/10.1332/174426413X662581>
- [6] Goetghebuer E, Loeys T. Beyond intention to treat. *Epidemiol Rev* 2002; 24: 85-90.  
<http://dx.doi.org/10.1093/epirev/24.1.85>
- [7] Ten Have TR, Normand SLT, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatr Ann* 2008; 38: 772-83.  
<http://dx.doi.org/10.3928/00485713-20081201-10>
- [8] Cartwright N. A philosopher's view of the long road from RCTs to effectiveness. *Lancet* 2011; 377: 1400-1.  
[http://dx.doi.org/10.1016/S0140-6736\(11\)60563-1](http://dx.doi.org/10.1016/S0140-6736(11)60563-1)
- [9] West SG, Duan N, Pequegnat W, Gaist P, Des Jarlais DC, Holtgrave D, Szapocznik J, Fishbein M, Rapkin B, Clatts M, Mullen PD. Alternatives to the randomized controlled trial. *Am J Public Health* 2008; 98: 1359-66.  
<http://dx.doi.org/10.2105/AJPH.2007.124446>
- [10] Hudgens G, Gilbert PB, Self SG. Endpoints in vaccine trials. *Stat Methods Med Res* 2004; 13: 1-26.  
<http://dx.doi.org/10.1191/0962280204sm356ra>
- [11] Dawid AP. Conditional independence in statistical theory. *J Roy Statist Soc B* 1979; 41: 1-31.
- [12] Dawid AP. The decision-theoretic approach to causal inference. In: Berzuini C, Dawid P, Bernardinelli L, editors. *Causality. Statistical perspectives and applications*. Chichester: Wiley 2012; pp. 25-42.  
<http://dx.doi.org/10.1002/9781119945710.ch4>
- [13] Begg C. Ruminations on the intent-to-treat principle. *Control Clin Trials* 2000; 21: 241-3.  
[http://dx.doi.org/10.1016/S0197-2456\(00\)00050-7](http://dx.doi.org/10.1016/S0197-2456(00)00050-7)
- [14] Armijo-Olivo S, Warren S, Magee, D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Phys Ther Rev* 2009; 14: 36-49.  
<http://dx.doi.org/10.1179/174328809X405928>
- [15] Robins JM, Greenland S. Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in and AIDS randomized trial. *J Am Stat Assoc* 1994; 89: 737-49.  
<http://dx.doi.org/10.1080/01621459.1994.10476807>
- [16] Altman D. Missing outcomes in randomized trials: addressing the dilemma. *Open Med* 2009; 3: 2.
- [17] National Research Council. *The prevention and treatment of missing data in clinical trials*. Washington: National Academies Press 2010.
- [18] Wright CC, Sim J. Intention-to-treat approach to data from randomized controlled trials: A sensitivity analysis. *J Clin Epidemiol* 2003; 56: 833-42.  
[http://dx.doi.org/10.1016/S0895-4356\(03\)00155-0](http://dx.doi.org/10.1016/S0895-4356(03)00155-0)
- [19] Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Stampfer MJ, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008; 19: 766-79.  
<http://dx.doi.org/10.1097/EDE.0b013e3181875e61>
- [20] Hernán MA, Robins JM. *Causal inference*. Boca Raton: Chapman & Hall/CRC 2013.
- [21] Weitkunat R, Lee PN, Baker G, Sponsiello-Wang Z, González-Zuloeta Ladd AM, Lüdicke F. A novel approach to assess the population health impact of introducing a Modified Risk Tobacco Product. *Regul Toxicol Pharmacol* 2015; 72: 87-93.  
<http://dx.doi.org/10.1016/j.yrtph.2015.03.011>
- [22] Rubin DB. Causal inference using potential outcomes: Design, modeling, decisions. *J Am Stat Assoc* 2005; 100: 322-31.  
<http://dx.doi.org/10.1198/016214504000001880>
- [23] Hoxby CM, Murarka S. Methods of assessing the achievement of students in charter schools. In: Berends M, Springer MG, Walberg HJ, editors. *Charter school outcomes*. New York: Lawrence Erlbaum 2008; pp. 7-38.
- [24] Feinman RD. Intention-to-treat. What is the question? *Nutr Metab* 2009; 6: 1.  
<http://dx.doi.org/10.1186/1743-7075-6-1>
- [25] Sackett DL, Wennberg JE. Choosing the best research design for each question. *BMJ* 1997; 315: 1636.  
<http://dx.doi.org/10.1136/bmj.315.7123.1636>
- [26] Welsh AW. Randomised controlled trials and clinical maternity care: moving on from intention-to-treat and other simplistic analyses of efficacy. *BMC Pregnancy Childbirth* 2013; 13: 15.  
<http://dx.doi.org/10.1186/1471-2393-13-15>
- [27] Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1989; 95(Suppl 2): 2-4.  
[http://dx.doi.org/10.1378/chest.95.2\\_Supplement.2S](http://dx.doi.org/10.1378/chest.95.2_Supplement.2S)
- [28] Shrier I, Steele RJ, Verhagen E, Herbert R, Riddell CA, Kaufman JS. Beyond intention to treat: What is the right question? *Clin Trials* 2014; 11: 28-37.  
<http://dx.doi.org/10.1177/1740774513504151>
- [29] Pocock SJ, Abdalla M. The hope and the hazards of using compliance data in randomized controlled trials. *Stat Med* 1998; 17: 303-17.  
[http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19980215\)17:3<303::AID-SIM764>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1097-0258(19980215)17:3<303::AID-SIM764>3.0.CO;2-0)
- [30] Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, Thabane M, Giangregorio L, Dennis B, Kosa D, Debono VB, Dillenburg R, Fruci V, Bawor M, Lee J, Wells G, Goldsmith CH. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol* 2013; 13: 92.  
<http://dx.doi.org/10.1186/1471-2288-13-92>