AAIM
Alliance for Academic Internal Medicine
Association of Professors of Medicine
Association of Program Directors in Internal Medicine
Association of Specialty Professors
Clerkship Directors in Internal Medicine
Administrators of Internal Medicine

AAIM Perspectives

AAIM is the largest academically focused specialty organization representing departments of internal medicine at medical schools and teaching hospitals in the United States and Canada. As a consortium of five organizations, AAIM represents department chairs and chiefs; clerkship, residency, and fellowship program directors; division chiefs; and academic and business administrators as well as other faculty and staff in departments of internal medicine and their divisions.

The Intention-to-Treat Analysis Is Not Always the Conservative Approach

Ian Shrier, MD, PhD,[a] Evert Verhagen, PhD,[b] Steven D. Stovitz, MD, MSc[c]

[a]Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Canada; [b]Amsterdam Collaboration on Health and Safety in Sports, IOC Research Centre for Prevention of Injury and Protection of Athlete Health, Department of Public and Occupational Health, EMGO+ Institute for Health and Care Research, VU University Medical Center, The Netherlands; [c]Department of Family Medicine and Community Health, University of Minnesota, Minneapolis.

The randomized trial design can be thought of as a means to answer 2 types of general questions: 1) what is the effect of assigning a treatment?; or 2) what is the effect of receiving a treatment?

In public health, we are normally concerned with the first question—the effect of assigning a treatment. If we implement a prevention or treatment program that is efficacious only under strict research conditions but people in the real world would not receive it for any possible reason, the program will not be effective. This real-world context is termed the "average causal effect" of assigning treatment and is best estimated by the intention-to-treat (ITT) analysis, where participants are analyzed according to the group to which they were assigned.[1-3] It has become the generally recommended approach for randomized trials.[4]

The ITT analysis estimates the average causal effect of assigning treatment to individuals, regardless of whether or not they adhered to the treatment. This represents a diluted effect when compared with what the effect of the treatment would have been had all participants adhered to their assigned treatment. In other words, the ITT is biased toward the null with respect to the average causal effect of the treatment in the population. Therefore, a different analysis is required if one wants to know the actual average causal effect of receiving a treatment, as opposed to the average causal effect of being assigned a treatment. In this article, we describe contexts when the ITT is not appropriate and why, and identify appropriate alternative analyses.

There are 2 reasons why the average causal effect of receiving a treatment may be more important than the ITT for some people. First, even in the public health domain, investigators may want to know what the average causal effect of a treatment program would be if they could improve participation in the program. For example, an ITT analysis might suggest that a vaccination treatment program was ineffective or minimally effective. However, in reality, the treatment program was very effective in the participants who received it, but the implementation of the program was inadequate. In this context, investigators may want to focus attention on improving adherence rather than on improving the vaccine. Also, the average causal effect of receiving a treatment is of primary interest to a patient deciding whether or not to take the treatment as recommended. This decision should be based on the probability of benefit if they take the treatment, or if they do not take

Funding: None.
Conflict of Interest: None.
Authorship: All authors took part in the preparation of the manuscript.

Requests for reprints should be addressed to Ian Shrier, MD, PhD, Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, 3755 Cote Ste-Catherine Road, Montreal, QC H3T 1E2, Canada.
E-mail address: ian.shrier@mcgill.ca

http://dx.doi.org/10.1016/j.amjmed.2017.03.023

the treatment. Providing a patient with information about the average causal effect of being assigned a treatment (ie, the ITT analysis) may not be very helpful, especially because such an average will vary widely among different implementation contexts even when the treatment effect is constant.

Because the ITT represents a diluted effect compared with the average causal effect of receiving a treatment, it is generally thought to represent a conservative approach, where "conservative" represents an approach that is less likely to accept a novel treatment as effective. For example, if the estimated treatment effect is less than the true treatment effect, the medical community is less likely to accept new effective treatments. In this article, we describe 3 scenarios where the ITT is not a conservative approach, even when missing data due to dropouts or other reasons are properly assessed.[2,3,5,6] Although some statisticians may be familiar with these concepts, our objective is to raise awareness among clinicians and clinician-researchers.

PERSPECTIVES VIEWPOINTS

- Clinicians need to be aware that intention-to-treat analysis sometimes overestimates the effect of receiving treatment.
- Intention-to-treat analysis may not be conservative when reference treatment is more effective than novel treatment.
- Intention-to-treat analysis may be biased away from the null when adherence rates are different between groups.
- Intention-to-treat analysis may be biased away from the null when treatment is effective in some patients but harmful in other patients.

First, if the reference (usual care) treatment is more effective than the novel treatment, then the approach is not conservative, even though there is bias toward the null. This scenario is commonly found in "noninferiority" trials (in both strictly controlled efficacy trials or pragmatic trials). Consider a reference group receiving proven usual care treatment by which blood pressure decreases by a mean of 10 mm Hg. In a novel treatment group, the mean blood pressure remains unchanged (ie, the novel treatment is ineffective). With $<100\%$ but similar adherence rates, the 2 groups will appear more similar; that is, biased toward the null. Investigators who conclude that the ineffective novel treatment is as effective as the usual care treatment might implement the novel treatment even though it was inferior. Thus, the ITT approach is not conservative in noninferiority trials, as it may result in accepting novel treatments that are less effective than reference/usual care treatments.

Second, bias may be away from the null if the adherence rates differ between study groups while both treatments are equally effective if taken.[2] Consider a problem that is universally fatal. When taken, usual care treatment results in survival of 50% of participants. Similarly, when taken, a novel treatment results in survival of 50% of participants. The treatments are thus equally effective. However, consider the context where the proportion of adherence in the novel treatment group is 100% and the proportion of adherence in the usual care treatment group is 50%. An ITT analysis of these data would suggest the novel treatment is twice as effective as the usual care treatment. Although the ITT analysis is correct with respect to the effect of assigning treatment to the entire population on average, individual patients who would adhere to the usual treatment should be aware that the 2 treatments are equally effective if taken as recommended. This is important for clinical care. For example, the usual care treatment might be much less expensive, and therefore, preferable for many patients. More generally, if all treatment options are equally effective, then the treatment associated with lowest adherence for any reason (eg, inconvenience, minor side effects, cost) will appear inferior if clinicians and patients are only provided an analysis based on an ITT approach.

In the above example, both treatments were considered equally effective for simplicity. Now consider a context where the novel treatment is superior to the usual care treatment, such that the novel treatment results in 60% survival and usual care treatment results in 40% survival (risk difference of 20%). If adherence is 100% in the novel treatment group, survival will be 60%. If adherence is 50% in the usual care group, survival will be only 20%. The ITT analysis now suggests the novel treatment will appear even more effective than usual care than it truly is (risk difference of 40% instead of 20%). Thus, the bias of ITT in this context would be away from the null.

In some contexts, the ITT analysis may give a qualitatively different answer and suggest that a treatment is worse than usual care when the novel treatment is actually superior. Consider the same survival rates (60% for novel treatment and 40% for usual care) where adherence rates are now 50% for the novel treatment and 100% for usual care. In this context, the ITT analysis would suggest that novel treatment is inferior (30% survival for the novel treatment group [50% adherence $\times$ 60% survival] and 40% for the usual care group) even though the novel treatment is more effective when it is received.

Finally, if neither treatment has any effect on the outcome, the ITT is generally considered unbiased.[2] However, there is the potential for bias away from the null compared with the average causal effect of

treatment when 1) a treatment is beneficial for some participants and harmful for others, and 2) the proportion of adherence is related to the outcome, such as when some of the side effects causing nonadherence are also markers for the outcome.[5,6] Consider the hypothetical trial presented in the **Table** that evaluates a treatment to reduce pain. Participant pairs 1 & 2, 3 & 4, and 5 & 6 are identical except for their assigned treatment. There are several different questions that might be asked. The ITT analysis provides the difference in results between participants assigned novel care vs usual care, regardless of which treatment they receive. For the effect of treatment itself, we might be interested in knowing the difference in results if every participant received novel care vs if every participant received usual care. This is known as the average causal effect of treatment (ACE). Alternatively, we might only be interested in the difference in results among those who would and could receive novel care vs those who would and could receive usual care, that is, exclude those participants who would never be able to receive a particular treatment because they have a contraindication, are not able to follow the intervention protocol for whatever reason, and so on. This is known as the complier average causal effect (CACE).[1] We might also be interested in learning if there are subgroups of participants who receive more benefit from the novel treatment vs usual care, and other subgroups who receive less benefit from the novel treatment vs usual care. Each of these questions is answered by a different analysis.

Columns 2 (pain level if receives usual care) and 3 (pain level if receives novel) illustrate the magnitude of pain measured on a scale from 0 (none) to 10 (maximum) that each participant would feel if they would receive the respective treatment. The mean pain level across all participants (Column 4) would be 5 whether all participants would receive usual care or all participants would receive novel treatment; the average causal effect is null. However, on an individual participant level, because the observed baseline score for each participant is 5, the novel treatment is beneficial for participants 1 & 2, harmful for participants 3 & 4, and equal for participants 5 & 6.

In an actual study, randomization would be expected to result in equal distribution of the 3 types of responders in each treatment-assigned group (Column 5). In real life, participants whose pain increases when given a treatment may switch, and the treatment received (Column 6) may be different from that assigned, for example, Participant #4 in the **Table**. Using the observed values in the study for pain and the treatment assignment, the ITT analysis estimates a treatment difference (Novel vs Usual Care) of −1.67. This suggests novel treatment is more effective (less pain) on average even though it is not (average causal effect = 0); therefore it is biased away from the null. Again, we used a pedagogical example where the average causal effect was null to better illustrate how the bias occurs. The same bias may be present if the average causal effect of the 2 treatments were different. For example, if the novel treatment actually increased pain mildly so that the mean pain level would be one point higher than baseline (6 units), the ITT analysis would suggest the novel treatment reduced pain by 0.67 units rather than actually increasing pain.

In some contexts, investigators may not be interested in the average causal effect of the entire population, but only in the subgroup of participants who would take the treatment assigned to them, that is, the CACE as described above. For example, some participants might avoid treatment because of an allergy. If we exclude these participants and all other participants who would never take one of the potentially assigned treatments, we are left with a subgroup of participants who would take their assigned treatment. This addresses a question some clinicians may ask: "Will the active treatment improve the health of my patients who would always follow my advice?" In the **Table**, we see that the ITT analysis yields an estimate that is between the CACE and the null, that is, the ITT may indeed be biased toward the null with respect to CACE, as shown in our example. In fact, with reasonable assumptions and CACE as the estimate of interest, McNamee[3] showed that the ITT is indeed always biased toward the null. Therefore, depending on the question of interest—the average causal effect in the population, or the average causal effect of only those that would follow recommendations—the ITT may or may not be biased toward the null.

How common is it for a treatment to be beneficial for some patients and harmful for others? We believe it is the norm for many conditions. For example, in rehabilitation, the same exercise program may stimulate healing in one patient but cause further damage in a different patient that is clinically similar. A concrete example would be a strengthening program for which individual participants have differing levels of expertise with technique. Alternatively, structural differences in bony alignment or biomechanical efficiencies may mean the strengthening program is optimal for one participant but excessive for another participant. More generally, any curative treatments with lethal side effects would satisfy the condition, such as airbags in cars that save adult lives but can be lethal for children. For quality of life, any treatment that is beneficial for the main outcome may still be associated with important inconveniences that result in an overall decreased quality of life. As such, investigators and clinicians need to carefully evaluate the underlying assumptions prior to accepting "rules of thumb," such as that the ITT is always a conservative approach.

There are 2 final points to emphasize. First, common "per protocol analyses" (analyzing only those

Table Potential Outcome and Observed Pain Scores for 6 Participants Randomized to Either Usual Care or a Novel Treatment, if Treatment May Be Both Beneficial and Harmful

	Potential Outcomes			Observed Outcomes			
Participant (Column 1)	Pain with Usual Care (Column 2)	Pain with Novel (Column 3)	Mean Counterfactual Pain (Column 4)	Treatment Assigned (Column 5)	Treatment Received (Column 6)	Observed Pain (Treatment Received) (Column 7)	Analyses: Novel vs Usual Care (Column 8)
1†	5 (NE)	0 (B)	Usual Care	Usual Care	Usual Care	5	ITT
2	5 (NE)	0 (B)	(6 * 5)/6 = 5	Novel	Novel	0	(0 + 5 + 5)/3 − (5 + 5 + 5)/3 = −1.67
3†	5 (NE)	10 (H)		Usual Care	Usual Care	5	ACE (from column 4)
4	5 (NE)	10 (H)		Novel	Usual Care	5	5 − 5 = 0
5†	5 (NE)	5 (NE)	Novel Care	Usual Care	Usual Care	5	CACE
6	5 (NE)	5 (NE)	(2 * 0 + 2 * 10 + 2 * 5)/6 = 5	Novel	Novel	5	[2*(0 − 5) + 2*(5 − 5)]/4 = −2.5

All participants have a baseline pain score of 5 prior to the study. Participants 1 and 2 are identical, 3 and 4 are identical, and 5 and 6 are identical. The pain score for all Usual Care is the same as baseline (no effect). The pain score for Novel treatment is reduced in participants 1 & 2 (Beneficial), increased for participants 3 & 4 (Harmful), and has no effect for participants 5 & 6 (No Effect). The average pain score for Novel treatment is the same as baseline (No Effect). Therefore, there is no difference in change in pain scores between Novel treatment and Usual Care, and the average causal effect is zero. †Participants who are considered in the Usual Care group for the intention-to-treat (ITT) analysis, which suggests Novel Treatment reduces pain more than usual care.

ACE = average causal effect, which is the effect in all patients (difference between columns 2 and 3) if everyone could and would receive both treatment alternatives; B = treatment is beneficial compared with baseline; CACE = complier average causal effect, which is the effect in the subgroup of patients (difference between columns 2 and 3) who will always follow their treatment assignment regardless of which treatment is assigned (participants 1, 2, 5 and 6); H = treatment is harmful compared with baseline; NE = treatment has no effect compared with baseline.

participants who actually followed treatment assignment) and "as treated" analyses (simply analyzing participants according to the treatment they received) are biased for any of the 3 questions we have asked and are strongly discouraged. This is because these analyses are valid only for the complier average causal effect if there were no differences in response to treatment between those who would and those who would not adhere to treatment assignment.[1] Although this assumption is valid for **Table** (participants 3 and 4 have the same response to Usual Care as the other participants) and both the per-protocol and as-treated analyses would provide the same answer as CACE, this assumption seems unrealistic in most contexts. Estimating the ACE and CACE requires other more advanced methods that are beyond the scope of this article. Second, investigators who are only interested in the effect of treatment assignment should still make individual participant data available, or report results stratified by assigned group and adherence status,[7] so that others can calculate the average causal effect or CACE should it be of greatest relevance to their own context.

In summary, the ITT analysis is extremely useful and answers questions such as "Will the active treatment improve health in the population if I fund the program?" or "Will changing treatment practices improve the health of my patients in general?" However, for many clinical questions about the effect of treatment received, clinicians and clinician-researchers should be aware that there are contexts where the ITT estimate is not a conservative approach and may be biased away from the null.

References

1. 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(1):28-37.
2. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48-55.
3. McNamee R. Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Stat Med*. 2009;28(21):2639-2652.
4. Fergusson D, Aaron SD, Guyatt G, Hebert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ*. 2002;325(7365):652-654.
5. Hernan MA, Robins JM. Measurement bias. Causal Inference. 2015:109-118. Available at: http://www.hsph.harvard.edu/faculty/miguel-hernan/causal-inference-book/. Accessed April 15, 2017.
6. Glymour MM, Nguyen QC, Matsouaka R, et al. Does Mother know best? Treatment adherence as a function of anticipated treatment benefit. *Epidemiology*. 2016;27(2):265-275.
7. Steele RJ, Shrier I, Kaufman JS, Platt RW. Simple estimation of patient-oriented effects from randomized trials: an open and shut CACE. *Am J Epidemiol*. 2015;182(6):557-566.