This is an author produced version of a paper published in The European Journal of Orthodontics. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the published paper:
Bondemark, Lars; Abdulraheem, Salem. (2018). Intention to treat (ITT) analysis as reported in orthodontic randomized controlled trials-evaluations of methodology and recommendations for the accurate use of ITT analysis and handling dropouts.. The European Journal of Orthodontics, vol. 40, issue 4, p. 409-413

URL: https://doi.org/10.1093/ejo/cjx084

Publisher: Oxford University Press

This document has been downloaded from MUEP (http://muep.mah.se).
Intention to treat (ITT) analysis as reported in orthodontic randomized controlled trials – evaluations of methodology and recommendations for the accurate use of ITT analysis and handling dropouts

Summary

Objective: To systematically evaluate in five orthodontic journals how many randomized controlled trials (RCTs) use intention to treat (ITT) analysis and to assess the methodological quality of the ITT analysis, and finally, to demonstrate in an academic way how outcomes can be affected when not implementing the ITT analysis.

Material and methods: A search of the database, Medline, was performed via PubMed for publication type “randomized controlled trial” published for each journal between 1 January 2013 and 30 April 2017. The five orthodontic journals assessed were the American Journal of Orthodontics and Dentofacial Orthopedics, Angle Orthodontics, European Journal of Orthodontics, Journal of Orthodontics and Orthodontics and Craniofacial Research. Two independent reviewers assessed each RCT to determine whether the trial reported an ITT or not or if a per-protocol analysis was accomplished.

Results: The initial search generated 137 possible trials. After applying the inclusion and exclusion criteria, 90 RCTs were included and assessed. Seventeen out of 90 RCTs (18.9%) either reported an ITT analysis in the text and/or supported the ITT by flow diagrams or tables. However, 6 RCTs applied and reported the ITT analysis correctly, while the majority performed a per-protocol analysis instead.

Conclusions: Nearly all the trials that applied the ITT analysis incorrectly, analysed the results using a per-protocol analysis, and thus, overestimating the results and/or having a reduced sample size which then could produce a diminished statistical power.
Introduction

In the healthcare sector, scientific assessment is undertaken to identify interventions which offer the greatest patient benefit while utilizing resources in the most effective way. To evaluate effectiveness, the randomized controlled trial (RCT) is the acknowledged standard and is considered to generate the highest level of evidence, followed by controlled trials (1, 2). The advantages of using the RCT methodology include the elimination of bias in treatment assignment, specifically selection bias (i.e. without therapists’ or patients’ preferences) and the confounding variables, hidden or out of control, will be of equal value in the groups. Even if RCT is the best way to minimize bias in ascertaining treatment effects, several types of bias can affect the external and internal validity of the trial: wrong randomization procedures, broken concealments, missing data or dropouts, non-compliance and the Hawthorne Effect are some of the biases that can occur in RCTs (1–5). The specific complications of missing data or dropouts and non-compliance are often of particular interest. One potential solution to the problem of missing data or a way of handling dropouts is to practice a statistical analysis called intention to treat (ITT) analysis. In accordance with Fisher et al., (6) the ITT analysis incorporates all randomized subjects in the groups to which they were randomly allocated, irrespective of their adherence with the entrance criteria, irrespective of the treatment the subjects actually received and irrespective of upcoming withdrawal from treatment or deviation from the protocol. Consequently, the ITT analysis is generally designated as “once randomized, always analysed” (7, 8), and therefore, disobey non-compliance protocol alterations, withdrawals and anything that may occur after randomization (9, 10). Moreover, ITT analysis avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers by accepting that non-compliance and protocol aberrations are likely to happen in certain clinical practices (10).
It can also be noted that the ITT analysis maintains the sample size. If non-compliant individuals or dropouts are eliminated from the conclusive analysis, i.e. a per-protocol analysis is performed, this will result in a significantly reduced sample size which then will produce a diminished statistical power (7).

It can be pointed out that, in the medical field, half of all the trials reveal that more than 10% of randomized patients have missing outcomes (11). Consequently, when multiple RCTs are pooled together, for example, in a Cochrane meta-analysis and with a small dropout rate for each trial, an important loss of outcome will still be evident in the compiled analysis if an ITT analysis is not applied. Nevertheless, care must always be taken to minimize missing responses and follow-up on those who withdraw from treatment (12, 13). Nevertheless, no consensus exists about how missing responses should be handled in the ITT analysis, and different approaches may be used in different situations. Thus, the ITT analysis appears to have different meanings for different authors or researchers (14).

Therefore, the primary aims of this study was to systematically evaluate how many RCTs use ITT analysis as reported in RCTs in five orthodontic journals and to assess the methodological quality of the ITT analyses. The secondary aims were to demonstrate in an academic way how outcomes can be affected by not using the ITT analysis and to give recommendations and advice to future researchers on how best to minimize and handle missing data and dropouts.

Material and methods

All reports of randomized controlled trials (RCTs) in five orthodontic journals were included. The journals were American Journal of Orthodontics and Dentofacial Orthopedics (AJODO), The Angle Orthodontist (AO), European Journal of Orthodontics (EJO), Journal of
Orthodontics (JO) and Orthodontics and Craniofacial Research (OCR). All except the Angle Orthodontist have applied the CONSORT statement (CONSORT 2010) (15).

The inclusion criteria were RCTs in humans and RCTs using the individual as the unit in one group and not included in other groups. Thus, RCTs applying the split-mouth design were excluded.

A search of the database, Medline, via PubMed was conducted for publication type “randomized controlled trial” published for each journal between 1 January 2013 and 30 April 2017. The journals were also hand-searched for the actual search period to cross check against the database search.

Two reviewers selected and categorized each article and then extracted the data independently. The number of inter-examiner agreements and disagreements were noted and calculated for the assessment of agreement. Then, all the disagreements were discussed and resolved by discussion to reach consensus.

If randomization was evident from the title and abstract of the articles, the full text of the study was read to verify and clarify the study design according to human RCTs applying the individual as unit; this was also to determine whether the study reported an ITT analysis or not or whether a per-protocol analysis was accomplished or not. Consequently, assessments were made according to deviations from correct ITT analysis, i.e. inadequately described and inadequately applied. Moreover, registrations were performed regarding deviations from random allocation and false inclusions as well as if insufficient or lacking dropout rates and descriptions were evident. Thus, the ITT analysis was assessed for each RCT by evaluating the number of participants before the start or at baseline and then at every outcome analysis occasions according to descriptions either in the text or in the Flow diagram of the article. If dropouts were evident, it was checked whether the dropouts were incorporated in the result
analysis (ITT) or if the dropouts were removed or not considered in the result analysis, i.e. a per-protocol analysis was performed.”

**Statistical analysis**

Fisher’s exact test was used to determine differences in categorical data. Differences with probabilities of less than 5% ($P$-value $< 0.05$) were considered to be statistically significant.

Inter-examiner agreement between the reviewers in assessing the articles in this survey with respect to title, abstract and full-text screening stage as well as for dichotomous and categorical variables during data extraction was assessed using Kappa score (16). Kappa values over 0.75 were considered excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor (17).

**Results**

The initial screening yielded 137 potentially relevant articles. Based on inclusion and exclusion criteria, 47 articles were excluded (nine of split-mouth design, four in-vitro studies, three prospective controlled and one retrospective controlled study). Thus, a total of 90 RCTs were included and assessed. The number of RCTs per journal is presented in Table 1, and the inter-examiner agreement was excellent (Kappa score 0.84).

Overall, 18.9% (17 of 90) of the RCTs either reported ITT analysis in the text and/or supported the ITT analysis by flow diagrams or tables that presented the dropouts in the final analysis. In contrast, 61.1% (55 of 90) of the RCTs either reported per-protocol analysis in the text and/or supported per-protocol analysis by flow diagrams or tables that presented the
dropouts in the final analysis (Table 1). Lastly, 20.0% (18 of 90) of the RCTs had not reported the number of dropouts at all (Table1).

The majority of the RCTs that performed ITT analysis were conducted in Europe (15 out of 17), more specifically, seven were from Sweden, five from the United Kingdom, and one each from Greece, Italy and Switzerland. Lastly, one RCT came from Egypt, and one from USA.

There were fewer RCTs that applied the ITT analysis in the Angle Orthodontist, although this was not statistically significant ($P = 0.056$).

Of the 17 RCTs that reported ITT analysis, six applied and reported the ITT analysis correctly and three of them did not have any dropouts. However, eight RCTs used the ITT analysis incorrectly and the majority performed a per-protocol instead of an ITT analysis, (Table 2). Finally, three RCTs performed the ITT analysis in a questionable way regarding conflict of statements, as the ITT analyses were stated or analysed in the results and discussion sections, but in the tables or flow diagrams, per-protocol analysis was evident.

**Discussion**

In this survey, we aimed to systematically evaluate how many RCTs used the ITT analysis and to assess the methodological quality of the ITT analysis; this was done with the intention to refrain from pointing out any individual study or authors who did not follow the ITT concept. Accordingly, we advocate learning by failure without pointing out anyone specifically responsible for any failure. Therefore, all RCTs in this survey were kept anonymous as a consideration for the authors who had conducted the trials. Nevertheless, it can be concluded that most of the RCTs that had used ITT analysis were from Europe (mainly Sweden and United Kingdom), and few were from other parts of the world. Without any solid explanation as to why most RCTs who report ITT are from Sweden and United Kingdom, it is
clear that researchers from these countries understand that it may be important to adopt an
ITT analysis.

The most important findings of this survey were that only 18.9% of the RCTs applied an ITT
analysis, and surprisingly, only a third of these used a correct ITT analysis. Instead, nearly all
the studies that applied the ITT analysis incorrectly analysed the results using the per-protocol
principle (Table 2). Thus, this evaluation indicated insufficient knowledge among orthodontic
researchers of how an ITT analysis should be performed. Furthermore, despite that four out of
the five journals (AJODO, EJO, JO and OCR) included in this survey require authors to
follow the CONSORT guidelines to publish an RCT and given that ITT analysis is one of the
CONSORT guidelines keys (15), it can be stated that the CONSORT guidelines are still not
fully complied with by the scientific journals.

Our findings of inadequate description and application of the ITT analysis are supported by
Lempesi et al. (2014) (4), who investigated the reporting quality of RCTs published in four
main orthodontic journals using the CONSORT statement checklist and concluded that the
reporting quality was suboptimal in various CONSORT areas. Consequently, 11.7% of the
RCTs reported ITT analyses adequately, and 85.9% had no description about ITT (4).
However, it is not just in the field of orthodontics that ITT analyses are deficient. In a survey
of RCTs published in four large medical journals (BMJ, Lancet, JAMA and New England
Journal of Medicine), it was found that the ITT analysis was often inadequately described and
applied (14).

Five recognized orthodontic journals were included in the survey, and four of them require the
authors to follow the CONSORT guidelines. Of course, additional orthodontic journals could
have been included in this survey, but the purpose of this survey was to provide an overview
of how ITT analysis was reported in RCTs published in well recognized orthodontic journals
with high degree of citations and a comparatively high impact factors. In fact, 3 of the
journals (EJO, AJODO and AO) have the highest impact factors and citations among existing orthodontic journals. Furthermore, these 3 journals contributed more than 90% of the RCTs included in this survey. Thus, it is unlikely that contributions from RCTs from additional journals would have changed our conclusion that many ITT analysis reported in orthodontic RCTs are suboptimal.

A tendency for the American Journal of Orthodontics and Dentofacial Orthopedics (AJODO) and the European Journal of Orthodontics (EJO) to publish a higher amount of RCTs with ITT analysis than the Angle Orthodontist (AO) was found. A possible explanation for this could be that, when submitting an RCT, both AJODO and EJO require authors to follow the CONSORT guidelines, which is not required by AO. Another possible explanation could be that the AO has strict word limitation which limits the authors to write not more than 3500 words, and thereby, a conceivable reason for missing many important items of the CONSORT 2010 statement in the submitted paper to the journal.

The time span for this survey was four years and four months, which can be considered relatively short, but the likelihood of finding higher levels of reported ITT analysis among RCTs produced earlier than 2013 is not imminent.

**Advice on how to best handle missing data or dropouts**

We recognised a misunderstanding of the ITT definition between authors through analysing and how they handled the dropouts. Handling dropouts can be tricky after randomization and before treatment. Therefore, it is always advantageous to know the reason behind the dropout, and this should be described in, for example, a flow diagram, i.e. that the subjects have moved from the area, regret participating in the trial, etc. Nevertheless, it can be noted that in multiple studies, if the dropouts occurred after randomization but before the intervention started, the patients were not considered as dropouts and not included in the final analysis,
which is actually defined as modified intention to treat (mITT) and not a pure ITT (18). In accordance with Fisher et al. (1990) (6), the ITT analysis incorporates all randomized subjects in the groups to which they were randomly allocated irrespective of their adherence to the entrance criteria, irrespective of the treatment the subjects actually received, and irrespective of any future withdrawal from treatment or deviation from the protocol. This indicates that multiple types of dropout analysis could be present: ITT, mITT, per-protocol and unclear analysis. A recent meta-epidemiological study investigated whether mITT is associated with biased estimates of treatment effects compared to ITT in rheumatoid arthritis RCTs. Seventy-two RCTs were included and analysed (23,842 patients), with 37 RCTs in the mITT group and 30 in the ITT group. They found no difference in the treatment effects between RCTs using mITT or ITT analyses populations (18), however, these results were in regard to rheumatoid arthritis studies and may not be appropriate for orthodontic studies.

When we exclude patients with missing data and use the per-protocol analysis, there is an obvious risk of producing a false positive results. The false positive results can easily be described by the following example, “Differences in treatment outcome for expansion plate with and without the ITT analysis”: Twenty-five patients were treated with expansion plate to correct unilateral posterior crossbite, and five dropped out because of cooperation failure. Thus, the success rate was 80%, i.e. 20 of 25 patients were successfully treated and the average maxillary expansion for the dropouts was counted as zero implying a mean maxillary expansion of 3.0 mm for the 25 subjects. If a per-protocol analysis had been applied instead, whereby the five dropouts would be excluded, this would result in a success rate of 20 out of 20 (100%) and a mean maxillary expansion of 3.7 mm. Both the success rate and the maxillary mean expansion will be significantly higher, $p = 0.041$ and $p = 0.024$, indicating that significant false positive results are evident if per-protocol was applied instead of an ITT analysis (19).
When long-term follow-up RCTs are performed, missing data and dropouts are usually unpreventable. Missing data may lead to bias, but one should keep in mind that an ITT analysis works by assumptions, which is also a type of bias (13). Unfortunately, there is no universal methodological approach for handling missing values. Multiple methods are available to deal with dropouts and these methods can be divided into two categories, simple imputation and multiple imputation methods. One example of simple imputation is Last Observation Carried Forward (LOCF). For multiple imputation category there are selection models and pattern mixture models as an example. In spite of this, some rules should be considered (13, 20); for example, the LOCF method works by using last measured values or data and impute those in all upcoming scheduled follow-up appointments if the patient has dropped out. Although this method has been criticized for bias introduction, we will nevertheless give two different orthodontic scenarios about LOCF. One scenario is regarding long-term follow-up of three retention methods, and the other considers a follow-up of white spot lesions (WSL) after treatment with fixed appliance. In the first scenario, the primary aim was to evaluate the Little Irregularity Index directly after the fixed orthodontic appliance was removed in addition to 1, 2 and 5 years after the fixed orthodontic appliance was removed. If dropouts occur at the 5-year follow-up registration and the LOCF method is applied, the registrations for the 2-year follow-up have to be used for those subjects who dropped out at the 5-year registration. However, it is well known from previous studies regarding incisor stability that the Little Irregularity Index worsens in the long term (21, 22). Consequently, when the LOCF procedure is applied, the risk is clear that the kept Little Irregularity Index at the 2-year follow-up registration (now also used as the 5-year registration) will be lower than it would have been if the subject had been maintained at the trial registration as planned at the 5-year follow-up. Accordingly, the LOCF procedure described in this scenario may result in false positive results.
The other scenario considers trials that evaluate WSL directly after removal of the fixed appliance and then at a follow-up one year after removal. The WSL will be registered in all patients directly after removal of the fixed appliance and at the 1-year follow-up. At the 1-year follow-up, some patients will have dropped out, and if the LOCF method is utilized, the same registrations of WSL at removal of the fixed appliance are also now used at the 1-year follow-up registrations for the dropout patients. Given that it is known from previous trials that WSL may regress due to various clinical technologies including no treatment (23), it can be assumed that the amount of WSL is higher at removal than at one year after removal of the fixed appliance. Thus, in this scenario, it is clear that keeping the outcome after treatment and using this outcome for the 1-year follow-up may result in false negative results.

From these two examples, it can be concluded that the compensation approach or handling of missing data is complicated and not a straightforward process, and thus, it is important to understand that it can induce different biases which are dependent on the type of outcome investigated and how different outcomes can develop during time.

**Recommendations**

When planning a new RCT, it is useful and recommended to investigate the pattern and timing of missing data in similar, previously performed RCTs. By identifying the pattern from previous RCTs, new ones can be designed in a different or appropriate way to avoid or decrease the missing data/dropouts as much as possible. For example, this could be accomplished by using a shorter trial period, or if longer registration times are at hand, using closer registration intervals (13).

At the stage when an RCT is designed:

- Start searching for any previously performed RCTs with investigations of outcomes within or tangent to the topic in question, and analyse the reasons and timing of
missing data and dropouts. Use this information to adjust your study for registration intervals or the length of follow-up, etc.

- Always use an ITT analysis and specify the method within the ITT analysis that you will use to handle the missing data (13).

At the publishing stage:

- State that you used an ITT analysis in the flow diagram, the methods, the results and in the discussion sections. Also, explain if a method within an ITT analysis has been performed to deal with the missing data, for example, the LOCF procedure.
- If an ITT analysis has not been applied, clearly state that a per-protocol analysis has been carried out.
- Document when the dropouts occurred and the reasons behind them (the CONSORT guidelines flow diagram or a graphical summary can be used).
- Analyse the background and the characteristics of the patients who have dropped out.
- If a per-protocol analysis must be applied, it can be useful in the discussion section to compare the results from a per-protocol and an ITT analysis and comment on how the different analyses may influence the results of the trial.

Conclusions

- Only 18.9% of the RCTs applied an ITT analysis, and surprisingly, only a third of these used a correct ITT analysis.
- Nearly all the trials that applied the ITT analysis incorrectly, analysed the results using a per-protocol analysis, and thus, overestimating the results and/or having a reduced sample size which then could produce a diminished statistical power.
If RCTs are to provide unbiased assessments, researchers must be aware of the accurate and correct usage of the ITT analysis.

- Orthodontic journals must also be aware that many ITT analyses reported in RCTs are suboptimal despite researchers being required to comply with the CONSORT guidelines.

**Funding**

This study was funded through the authors’ academic institution.

**Conflict of Interest**

None.

**References**


Table 1. Number of RCTs per journal and number of RCTs reporting intention to treat (ITT) analysis, per-protocol (PP) analysis, and number of RCTs that not mentioned or considered dropouts.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Number of RCTs per journal (%)</th>
<th>Number reporting ITT of the total RCTs (%)</th>
<th>Number reporting PP of the total RCTs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AJODO</td>
<td>36 (40.0)</td>
<td>8 (8.9)</td>
<td>24 (26.7)</td>
</tr>
<tr>
<td>AO</td>
<td>24 (26.7)</td>
<td>1 (1.1)</td>
<td>12 (13.3)</td>
</tr>
<tr>
<td>EJO</td>
<td>22 (24.4)</td>
<td>7 (7.8)</td>
<td>13 (14.4)</td>
</tr>
<tr>
<td>JO</td>
<td>5 (5.6)</td>
<td>0 (0.0)</td>
<td>4 (4.4)</td>
</tr>
<tr>
<td>OCR</td>
<td>3 (3.3)</td>
<td>1 (1.1)</td>
<td>2 (2.2)</td>
</tr>
<tr>
<td>Total</td>
<td>90 (100)</td>
<td>17 (18.9)</td>
<td>55 (61.1)</td>
</tr>
</tbody>
</table>

Table 2. The eight RCTs that reported using the ITT analysis but applied it inadequately.

Descriptions of the inadequate applications:

- One thousand patients were randomized, 826 were analysed at T1 and 684 at T2, and the analysis at T2 also went under power, i.e. a per-protocol, not an ITT analysis, was performed.

- Eighty-one patients were randomized and 67 were analysed, i.e. a per-protocol analysis.

- One patient missed all the follow-up appointments and was considered a dropout but included in the main outcome analysis. Three patients showed up for the 2-month follow-up and then disappeared, one patient came back with a broken retainer to the emergency clinic after the trial was ended and was considered as a failure, and the other two patients were considered successes since they had not returned to the emergency clinic. Patients who did not show up should have been considered failures, i.e. that the retainer did not survive.

- Eighty-one patients were randomized and 62 were analysed, i.e. a per-protocol analysis.

- One hundred and thirty-five patients were randomized and 115 were analysed, i.e. a per-protocol analysis.

- Forty patients were randomized, and five patients dropped out. Forty patients were analysed regarding cost effectiveness analysis and 35 patients for maxillary expansion, i.e. an ITT analysis was made for the cost effectiveness, but a per-protocol analysis was conducted for the treatment effects.

- Sixty-two patients were randomized, and one patient dropped out at T1 and five patients at T2. Sixty-two patients were analysed at T1 and 57 at T2, i.e. a per-protocol analysis was performed.

- Sixty-one patients were randomized, six did not start the intervention, nine patients dropped out after the intervention started, and 46 patients were analysed, i.e. a per-protocol analysis was performed instead of an ITT analysis.