International Journal of Epidemiology
© International Epidemiological Association 1992

Vol. 21, No. 5
Printed in Great Britain

Intention-to-Treat Analysis: Implications for Quantitative and Qualitative Research

DAVID J NEWELL

Newell D J (Department of Medical Statistics, University of Newcastle upon Tyne, UK). Intention-to-treat analysis: implications for quantitative and qualitative research. *International Journal of Epidemiology* 1992, **21**: 837–841
Intention-to-treat analysis is an important aspect of randomized controlled trials of health care interventions. The concept is now widely accepted in theory, but not always implemented in practice. Failure to analyse by intention-to-treat can give misleading and indeed life-threatening interpretations. In some studies, a case is put for estimating the effect that would have been observed if all patients had received the allocated treatment. Situations where this is valid are rare, but an example is given of such an exceptional study. The relevance of the intention-to-treat concept is not always taken into account in qualitative research. Interviews with new mothers who delivered their babies at home in a hypothetical controlled trial of home versus hospital confinement would provide fascinating accounts of the pleasures of successful delivery at home. But by definition the interviews would exclude the hazard and drama of necessary transfers to hospital due to complications in late pregnancy and early labour. The intention-to-treat approach would avoid this bias.

The expression 'intention-to-treat' analysis for randomized controlled trials probably first appeared in print in the 1961 edition of Bradford Hill's *Principles of Medical Statistics*,[1] although the concept appears earlier. Essentially, it says that all patients randomly allocated to one of the treatments in a trial should be analysed together as representing that treatment, whether or not they completed, or indeed received that treatment. The idea was clarified and amplified by Schwartz and Lellouch,[2] and further codified in their 1980 book.[3] Although careful in their approach, these authors left the impression that analyses other than 'intention-to-treat' were equally valid in some circumstances. The purpose of randomization is to avoid selection bias and to generate groups which are comparable to each other. Any changes to these groups by removing some individuals' records or transferring them to another group destroys that comparability. The object of this paper is to emphasize that intention-to-treat can only be disregarded in rare circumstances in health care research, with important implications for qualitative as well as quantitative research.

My own introduction to the importance of intention-to-treat analysis and the possible dangerous misinterpretations from using other methods came in a 1962 trial of low temperature incubators for premature babies.[4] Babies were randomly allocated to a temperature of 85°F or 98°F (say 30°C or 37°C). Halfway through the study, while the chief investigator was overseas, someone undertook an unauthorized interim analysis and stopped the study. Admittedly, not much was known then about the effect of interim analyses on significance levels,[5] but worse, that analysis also took no account of the recent but already established intention-to-treat principle. It concentrated on the mortality of the babies who *maintained* a temperature of 85°F, and found that they had a remarkably low death rate. It was not noticed that those allocated to the low temperature but who did not maintain it (the non-compliers in today's nomenclature) had a high death rate. On the return of the chief investigator, a proper intention-to-treat analysis of the interim data showed no such effect and the study was resumed. At its end, it was concluded that

Department of Medical Statistics, University of Newcastle upon Tyne, UK
Present address: Centre for Health Economics Research and Evaluation (University of Sydney), Westmead Hospital, Westmead, NSW 2145, Australia

low temperature incubators were detrimental, and they were abandoned. It is a worrying thought that had the interim analysis been accepted by the eminent paediatrician investigator and published under his name, other cold incubators would have been introduced, and other babies would have died unnecessarily. Intention-to-treat analysis can save lives!

RANDOMIZED CONTROLLED TRIALS

For any RCT of a health care intervention (e.g. formal rehabilitation compared with none, day case surgery compared with inpatient care, two different recruitment methods for mammography), the broad research outline is as shown in Figure 1.

The first step, patient selection, comprises inclusion and exclusion criteria including patient consent. That done, randomization to Treatment A or Treatment B should take place 'as late as possible', when diagnostic and consent procedures have all been completed. This helps to reduce non-compliance to its reasonable minimum.

I once worked with a surgeon who took this advice so literally that he waited until the patient's abdomen was open and the diagnosis confirmed before randomizing. He had a silver coin sterilized along with the scalpels, and tossed it right there in the operating theatre. If if fell heads, he would cut out one foot of gut; if tails, cut out 10 feet. That may sound a bit extreme, but the anecdote is included to emphasize the requirement to randomize as late as possible.

In intention-to-treat analysis, the randomization not only decides the allocated treatment, it decides there and then how that patient's data will be analysed, whether or not the patient actually receives the prescribed treatment.

Objections have been raised to this method of analysis. Suppose, for dramatic simplicity, that patients are randomly allocated to medical or surgical treatment. Those allocated to medical treatment are given medication immediately, while those allocated to surgery may require preparation, possibly waiting a few days or weeks for an available surgical theatre time-slot. If a patient should happen to die before reaching the operating theatre, a surgeon might be inclined to say 'that death should not count against the surgical option—I didn't get a chance to put my knife into the patient'. The physician would rightly claim that if the surgeon could discount these—obviously the sickest—patients, the comparison would not be fair. In fact, the surgical *programme* includes some (inevitable) delays, and all mortality occurring after the decision to perform surgery must correctly be assigned as part of the outcome of that programme.

However, other methods of analysis have been proposed which take account of whether the allocated treatment was actually received. Thus, referring to Figure 1, by the end of the trial there are four groups of patients: (1) those allocated to A who did not complete A; (2) those allocated to A who did complete it; (3) those allocated to B who completed it, and (4) those allocated to B who did not complete it.

Intention-to-Treat Analysis (otherwise known as 'Pragmatic Trial' or 'Programme Effectiveness Analysis') compares 1 + 2 with 3 + 4. Efficacy Analysis (otherwise known as 'Explanatory Trial' or 'Test of Biological Efficacy') compares 2 with 3 ignoring 1 and 4. Treatment Received Analysis (otherwise

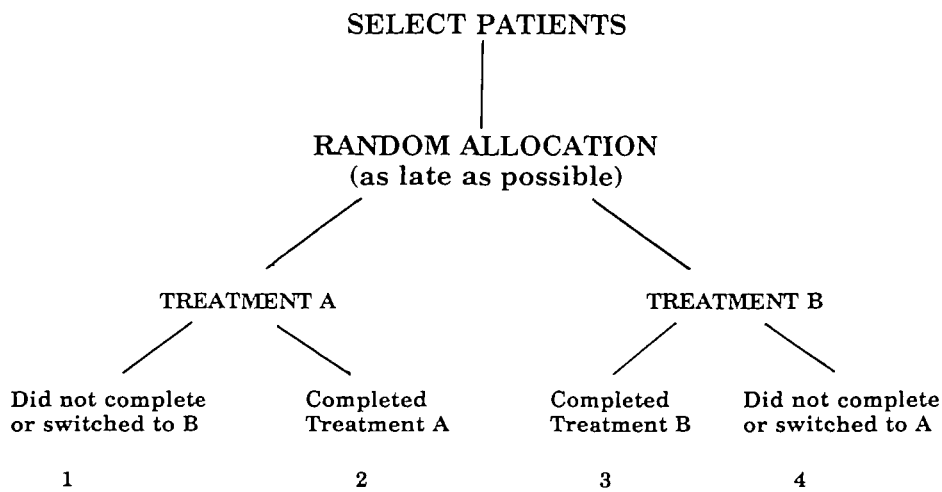

FIGURE 1 *A simplified schema for a randomized controlled trial*

known as 'As treated'*) compares 1 + 3 with 2 + 4 when treatments are switched.

EXAMPLE OF METHODS OF ANALYSIS

The two-year follow-up data of one of the early trials of Coronary Artery Bypass Surgery (CABS)[6,7] provide a good example of the results of these three methods of analysis. Before considering these data, it should be noted that the approach of the CABS research team was impeccable. They refused to do an 'as treated' analysis: 'We have refrained from comparing all patients actually operated on with all not operated on: this does *not* provide a measure of the value of surgery'.[7] Although they did publish an 'efficacy' analysis, those figures 'were compared merely to demonstrate the extent of deviation of such results from those obtained *by the proper method of comparison*'. Conclusions drawn from longer follow-up[7] were different from the two-year results, but they are irrelevant to the purposes of this paper.

Table 1 shows the two-year mortality data. Of the 373 allocated to Medical Treatment, 50 switched to surgical treatment, and they did quite well, with only two deaths. Of 395 allocated to Surgery, 26 received only medical treatment: of them six died (all of them in fact before the planned operation date).

TABLE 1 *Numbers of survivors and deaths, 2 years after allocation to CABS or medical treatment*

	Allocated to medicine		Allocated to surgery	
	Received surgery	Received medicine	Received surgery	Received medicine
Survived 2 years	48	296	354	20
Died	2	27	15	6
Total	50	323	369	26

Table 2 shows the results of analysing these results by the three methods. Intention-to-treat analysis (as used, correctly, by the research team) found a 7.8% mortality in those allocated to medical treatment, and a 5.3% mortality in those allocated to surgery. The difference between these two rates could easily have arisen by chance if the two treatments were equally effective (two-tailed $P = 0.17$). If, however, the analysis had been restricted to those (the 'compliers') who actually received the treatment allocated to them, the two mortality rates (8.4% for medicine, 4.1% for surgery) would have been significantly different ($P = 0.018$) even allowing for 3 or 4 interim analyses. Finally, an 'As treated' analysis would have wildly exaggerated the apparent value of surgery, suggesting that medical treatment has a 9.5% mortality rate compared with 4.1% ($P = 0.003$) for surgery.

TABLE 2 *Two-year mortality rates as calculated by three methods*

	Allocated to			
Analysis by	Medicine	Surgery	χ^2_1	P
Intention to treat	29/373 (7.8%)	21/395 (5.3%)	1.9	0.17
Compliers only	27/323 (8.4%)	15/369 (4.1%)	5.6	0.018
As treated	33/349 (9.5%)	17/419 (4.1%)	9.1	0.003

EXCEPTIONS TO THE INTENTION-TO-TREAT IMPERATIVE

The main message of this paper is that in any test of a public health intervention, an intention-to-treat analysis should be undertaken. This message applies both to randomized controlled trials and to cohort studies of a quasi-experimental design. While this paper was in preparation, an excellent account of the intention-to-treat principle and postulated alternative analyses was published by Lee *et al.*[8] Although the details of their example and their analytical approach were somewhat different, their conclusions were essentially the same: valid analyses should be based on intention-to-treat.

There are, however, exceptions to any rule. In some instances, even the hardest work may leave some cases 'lost to follow-up', so that the required endpoint cannot be identified (e.g. alive/dead at a specified date or anniversary of randomization). If such cases are few, sensitivity analysis assuming they are all alive or they are all dead may, with luck, leave the major conclusion unchanged, with rather similar point and interval estimates of treatment effect. Otherwise, it may be that some of the characteristics of non-compliers are known (e.g. socioeconomic status), and that the outcome variable has a generally known relationship to that characteristic. Again, with luck, an efficacy result might be projected to an 'intention-to-treat' result by an *a fortiori* argument.

Finally, it might be possible to model the conditions of a particular trial and obtain valid estimates for at least a subset of the patients. A recent interesting example follows.

*In the conference presentation on which this paper is based, this method was succinctly described as 'Garbage Analysis'. In deference to colleagues who have used the method, the less pejorative term 'As treated' is used in the text.

EXAMPLE OF AN EXCEPTION

A randomized controlled trial[9] compared some Indonesian children given Vitamin A tablets with a control group who under local law could not be given a placebo. A straightforward intention-to-treat analysis showed a control death rate of 74/11 588 = 6.4 per 1000. Those allocated to Vitamin A had a significantly lower rate of 46/12 094 = 3.8 per 1000, even though some 20% of the children had not complied with taking the allocated tablets, for very good reasons. In general, in the group allocated to Vitamin A the non-compliers were sicker than the compliers. The report indicates that Vitamin A tablets were not otherwise available, so there were no difficulties of non-compliance in the control group.

In a subsequent analysis,[10] the authors asked what would be the efficacy of a similar amount of vitamin introduced compulsorily into a staple food, which everyone would eat (zero non-compliance). Their approach was to postulate:

i) that the whole population could be divided into the sorts of children who would comply and would not comply when given any tablet.

ii) that the proportion of non-compliers observed in those allocated vitamin treatment (about 20%) would have applied to the controls if they had been allocated placebo tablets.

iii) that the mortality rate in the non-compliers for placebo would be the same as for non-compliers for vitamin (as they received no treatment). This rate was observed as 34/2419 = 14.1 per 1000.

From these assumptions, the number of the control group deaths that would have been amongst non-compliers can be calculated by simple proportions. Subtraction then yields the number of children and the number of deaths in the potential compliers. These worked out to 9270 and 41 respectively, giving a death rate of 4.5 per 1000.

Finally, this rate in the control potential compliers can be compared with the observed death rate in the vitamin group's compliers, 12/9675 = 1.2 per 1000.

The relative risk of death after taking Vitamin A is thus 28% (which the authors show has a 95% confidence interval [CI] of 13–59), in the complier type. No estimate is possible of relative risk in the non-complier type. Nevertheless, a valid estimate of relative risk for 80% of the population is well worth having.

Valuable in itself, this particular example illustrates a class of problems where modelling of compliance is possible and realistic. Note that the valid result is quite different from that which would be found from either a 'compliers only' or an 'as treated' analysis. This rather unusual analysis was possible only because the authors could postulate a situation in which dosing the food-supply would ensure that there were no non-compliers.

IMPLICATIONS OF 'INTENTION-TO-TREAT' ANALYSIS FOR QUALITATIVE RESEARCH.

The earlier analysis of CABS and Table 1 shows that results can be dramatically wrong if a randomized controlled trial is analysed on anything other than an 'intention-to-treat' basis. For the qualitative researcher, the implication is that great care must be taken in selecting a sample to interview. In any health care intervention, some of the target clients cannot or will not 'comply'. In a study of the value of rehabilitation to assist clients to return to work after an industrial injury, some individuals allocated to a rehabilitation counsellor will return to work before the first counselling session. If the purpose of the qualitative research is to produce a generalizable, rather than a purely local result, the research question might well be framed 'What would be the effect on individuals of introducing such a programme in another similar locality?' For a complete answer, the sample interviewed should include some of those who went back to work without benefit of counselling, in the appropriate proportion. In other words, the sample should be selected on an 'intention-to-treat' basis.

Similar considerations would apply if the planned intervention was domiciliary midwifery. In an RCT, the only mothers who could be randomized would be indifferent as to whether they had home or institutional delivery. However, some of those randomly selected for home delivery might very well be transferred to a hospital because of clinical complications arising during pregnancy or labour which could not be adequately handled in the home setting. Again, if a domiciliary service were introduced elsewhere, due account would need to be taken of the need to provide for these 'non-compliers'. In a qualitative review of the service by interview of mothers on the domiciliary programme, those who had to transfer to hospital must be included. If they were excluded, interviews of those who successfully delivered at home might well present too rosy a picture of a domiciliary maternity service.

CONCLUSION

In any study of a health intervention, it is essential to remember patients or clients who would not or could not complete the planned intervention, and include them appropriately in the analysis.

REFERENCES

[1] Hill A Bradford. *Principles of Medical Statistics* 7th edn. London: The Lancet, 1961; p 259.

[2] Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chron Dis* 1967; **20**: 637–48.

[3] Schwartz D, Flamant R, Lellouch J. *Clinical Trials*. London: Academic Press, 1980.

[4] Jolly H, Molyneux P, Newell D J. A controlled study of the effect of temperature on premature babies. *J Pediatr* 1962; **60**: 889–94.

[5] Pocock S J. Interim analyses for randomised clinical trials. *Biometrics* 1982; **38**: 153–62.

[6] European Coronary Surgery Study Group. Coronary artery bypass surgery in stable angina pectoris: survival at two years. *Lancet* 1979; **i**: 889–93.

[7] European Coronary Surgery Study Group. Coronary artery bypass surgery: survival to five years. *Lancet* 1982; **ii**: 1173–80.

[8] Lee Y J, Ellenberg J H, Hirtz D G, Nelson K B. Analysis of clinical trials by treatment actually received: is it really an option? *Stat Med* 1991; **10**: 1595–1605.

[9] Sommer A *et al.* Impact of Vitamin A supplementation on childhood mortality: a randomised controlled community trial. *Lancet* 1986; **i**: 1169–73.

[10] Sommer A, Zeger S L. On estimating efficacy from clinical trials. *Stat Med* 1991; **10**: 45–52.