Nephrol Dial Transplant (2013) 28: 826–832
doi: 10.1093/ndt/gfs307

ndt
Nephrology Dialysis Transplantation

*Polar Views in Nephrology*

# Con: Randomized controlled trials (RCT) have failed in the study of dialysis methods

Zbylut J. Twardowski and Madhukar Misra

*Correspondence and offprint requests to*: Zbylut J. Twardowski; E-mail: twardowskiz@health.missouri.edu

Division of Nephrology, Department of Medicine, University of Missouri, Columbia, MO, USA

Keywords: hemodialysis, observational studies, randomized controlled trials, statistical differences, statistical errors

## ABSTRACT

All progress in dialysis methods was made in research presented in case reports, case–control studies and other observational studies. On the contrary, randomized controlled trials (RCTs) did not bring any valuable results. Comparison of the value of peritoneal dialysis and hemodialysis (HD) in RCTs was not completed because of recruitment problems. Four RCTs in HD did not provide any useful data. The worst example was the National Cooperative Dialysis Study, which committed a Type II statistical error rejecting the time of dialysis as an important factor determining the quality of dialysis. This study also provided the basis for the establishment of the *Kt/V* index as a measure of dialysis adequacy. This index was accepted by the HD community, having been established in a sacrosanct RCT, led to short dialysis, and possibly higher mortality in the USA. The second trial (the HEMO study) committed a Type III statistical error asking the wrong question and did not bring any valuable results, but at least it did not lead to deterioration of dialysis outcomes in the USA. The third, the Frequent Hemodialysis Network Trial Group, did not bring forth any valuable results, but at least confirmed what was already known. The fourth, the Frequent Hemodialysis Network Nocturnal Trial, committed a Type II statistical error because of tremendous recruitment problems leading to an inadequate number of subjects. Moreover, the study methodology was absolutely unreliable.

## INTRODUCTION

A randomized controlled trial (RCT) is considered the most powerful scientific tool available in the health care field used to ascertain an objective measure of the superiority of one treatment method over another. But does this also hold true when comparing dialysis methods? Before we review and critique the difficulties encountered in the RCTs of dialysis methods, let us review some inherent problems related to the assessment of differences.

## ASSESSMENT OF DIFFERENCES

The issue of statistical significance versus clinical importance was the subject of an excellent review by the late Feinstein in 1988 [1]. According to him, while assessing the 'significance' of difference, one has to be aware of the kind of difference under study: ranking, quantitative, stochastic, clinical (mortality and quality of life) and architectural (quality of comparison). We will not discuss ranking differences, which are used in competition and are always significant. Another difference, which also does not require statistical analysis, is quantitative difference [an apparently large difference that fulfills the criterion of the 'traumatic interocular test' (TIT): a difference so profound and obvious that, metaphorically, it hits one between the eyes. 'You don't need a fancy P-value or other statistics to say, 'Yes, that's a real difference'] [1]. Some such landmark examples include thyroxine for myxedema in 1891, insulin for diabetic ketoacidosis in 1922, vitamin $B_{12}$ for pernicious anemia in 1926, penicillin for G+ cocci sepsis in 1941, defibrillation for ventricular fibrillation in 1948 and imatinib for chronic myeloid leukemia in 2002 and gastrointestinal stromal tumors in 2005 [2].

The third type of difference is stochastic. The word stochastic, from the Greek word στοχαζεσθαι (*stokhazesthai—to guess*), indicates the idea of randomness, i.e. how chancy the difference is. This is the ordinary meaning of the term 'statistical significance.' Historically, different levels of

certainty were established to qualify whether the difference between the groups was real or incidental [3]. This concept was primarily studied by two major schools of thought in the 1920s and the 1930s [4]. The school of Fisher [5] propounded that if the difference between the means of two groups was small, then the groups were deemed to be from the same population with the same mean (null hypothesis). Fisher established that a P-value of $< 0.05$ (chance difference probability of <1–20) be considered significant and the null hypothesis should therefore be rejected. The second school, of Neyman and Pearson, introduced the concept of 'errors.' The error that incorrectly rejects the null hypothesis is called Type I or α error, whereas the error that incorrectly accepts the null hypothesis is called a Type II or β error [6]. This error is caused by insufficient sample size. Later, Kimball [7] postulated a Type III error, an error that gives the right answer to the wrong problem. A Type IV error was subsequently postulated as a type of error that solved the right problem too late [8].

By comparison of two groups of subjects, the higher the difference between the groups, the smaller the number of subjects needed to show a stochastic significance at the level of $P < 0.05$. Only 10 subjects may be needed in each group if the mortality in one group is 80 and 20% in the other group. On the other hand, a trivial difference in mortality such as between 37 and 36% between the two groups would require >9000 subjects in each group to prove that the difference of 1% is stochastically significant but still may be clinically irrelevant. 'The results are splendid for policy-making decisions of pharmaceutical companies and regulatory agencies.' However, they 'may not be pertinent to the important distinctions of pertinent clinical subgroups that must be considered when treatment is chosen for individual patients' [3].

## HOW TO DETERMINE THE STOCHASTIC DIFFERENCE?

There are many methods for determination of such a difference, often termed the 'rules,' 'levels' or "hierarchies.' An example of hierarchy of evidence was provided in a Harveian oration by Rawlins [2] (Table 1).

**Table 1. How to determine the significance of stochastic difference?**

| |
|---|
| Randomized controlled trial (RCT) |
| Observational studies |
| Historical controlled trials |
| Non-randomized, contemporaneous controlled trials |
| Case-controlled studies |
| Before-and-after designs |
| Case series and case reports |

### Randomized controlled trials

RCTs are considered the most reliable of such methods. In the early 1920s, a method for randomization of experimental studies was established by Sir Ronald Aylmer Fisher, a statistician at the Rothamsted Agricultural Experimental Station. The problem that was being studied was to compare the effects of different fertilizers on the yields of potatoes [9]. The old method was to apply each fertilizer to an entire field and compare yields between fields. Because some fields may be more fertile than others, Fisher divided the fields into rows and then into small plots within the rows, randomly assigning fertilizers to the plots before assessing the results for each fertilizer.

Nowadays, RCTs are considered 'a must' to establish the reliability of an observed difference. The RCT as performed by Fisher gave excellent results. However, it is open to question if the same can be said for its universal applicability in medical research, particularly when studying therapeutic methods or devices. This is due to the fact that there are inherent differences in the nature of the scientific question that was studied by Fisher and the RCTs in medicine as they apply to the study of methods. Potatoes and plots are different from patients, as nobody asks them to agree to be assigned to a particular row; there are no exclusions; they are not asked to do anything; they are always compliant; they do not withdraw from a trial; comparison is free of bias; no special equipment, venue or skills are needed for different plots. Extrapolating the need for randomization to all clinical hypothesis testing (or else the clinical equipoise would be violated) can create problems. Hill, who introduced RCTs in medicine [10], himself warned that 'any belief that the controlled trial is the only way would mean not that the pendulum had swung too far but that it had come right off its hook' [11].

Moreover, there is an important difference between the studies of pharmaceuticals and therapeutic methods. In the study of pharmaceuticals, one group takes one medication and another group a placebo or another medication. In the study of therapeutic methods, the patients are asked to perform procedures with different pieces of equipment and/or at different venues. This creates inherent problems with recruitment and compliance of subjects with such studies.

### Observational studies

Observational studies are often considered inferior to RCTs, but are they? All the quantitative differences fulfilling the criterion of TIT were achieved in historical controlled trials. In the middle of the 20th century, Doll and Hill performed case-controlled studies strongly suggesting that smoking was associated with lung cancer [12, 13]. Owing to a lack of randomization, they invoked criticism from Fisher [14]. Nevertheless, according to these case-controlled studies, cigarette smoking is considered harmful with many legislating mandates based on this association although it has never been confirmed in RCTs.

### Before-and-after designs

Excellent, valuable information can be obtained from trials using before-and-after designs. They require a small number of subjects and are quick to conduct, inexpensive and very reliable under certain predefined conditions. With such a design, patients are their own controls, so the patients know after the study what was good for them. It is obvious that such trials do not allow one to determine mortality differences and they are also not useful in conditions with a fluctuating natural history. However, dialysis patients are excellent candidates for a before-and-after design, because the course of their disease is very stable if the treatment remains unchanged. Their condition may slowly deteriorate because of a fall in urine output. If patients are selected for a trial after urine output is close to zero, then any change in patient condition is very likely related to the dialysis method. The fact that mortality differences cannot be determined in such studies is not really important as 'there is more to life than absence of death.' Substitutes such as blood pressure control, left ventricular mass and nutritional indices are excellent indicators of probability of survival. Some RCTs incapable of determining mortality differences because of inadequate samples have used similar surrogates.

## STATISTICAL METHODS IN DIALYSIS

### Observational studies

Let us now analyze the types of studies that have been performed in dialysis research and the usefulness of the results so obtained. Let us reverse the postulated hierarchy and start with case series and case reports. Table 2 shows studies fulfilling the criterion of TIT. All of them were case reports, case series or historical controlled trials. According to Scribner, the father of chronic hemodialysis (HD), 'Successful treatment of Clyde Shields represents one of the few instances in medicine where a single success was all that was required to validate a new therapy' [15]. None of these studies, forming the very edifice of dialysis therapy, was required to demonstrate statistical significance to prove its striking results. As yet another example, the frequency of HD (thrice a week) used in the majority of HD prescriptions was established in 1965 [16]. At that time, the total weekly dialysis time was over 20 h and the mortality in the USA was <10% per year.

### Before-and-after designs

Between March 1969 and May 1973, one of us (Z.J.T.), while living in Poland, carried out research to determine the optimal duration and frequency of dialysis that would eradicate all symptoms and signs of uremia and lead to full rehabilitation. This could then be considered adequate dialysis. Statistical evaluation by the before-and-after method using Student's paired *t*-test showed that more frequent and/or longer HDs improved the control of blood pressure, hematocrit, serum albumin and nerve conduction velocity. The results were so impressive that it was predicted that more frequent and long dialyses would soon be the standard of care [17]. This prediction proved to be wrong, not because the conclusions were wrong, but because the development of dialysis turned in another direction.

### Randomized controlled studies in dialysis

The major incentives to shorten dialysis were economical and organizational and there was a perceived need to dialyze all the patients during, at most, three hemeral shifts. The support for this approach came from the National Cooperative Dialysis Study (NCDS). This was an RCT which compared outcomes in patients assigned to high or low urea levels and short (199 or 194 min) or long dialysis (269 or 271 min) times. Time of dialysis was rejected as an important factor based on a P-value of 0.056 [18, 19]. This was caused by insufficient sample size. It was then widely accepted that urea clearance, but not dialysis time, is important for patient well-being. On the basis of this study, Gotch and Sargent developed the $Kt/V_{urea}$ index [20]. According to their assessment of the NCDS, a value of this index of 0.95–1.0 was sufficient to provide adequate dialysis. Combining '*K*' and '*t*' in the numerator indicated that shorter dialysis time may be compensated for by higher urea clearance. The NCDS was the first randomized controlled trial in dialysis, and it was subject to a Type II (β) error of accepting the wrong null hypothesis that time of dialysis is not important for quality of dialysis if compensated for by higher urea clearance. This undoubtedly led to rejection of the importance of dialysis time and may have contributed to poor outcomes on HD in the USA. As more efficient dialyzers were developed, the length of dialysis decreased to three and even to 2 h. Annual mortality of HD patients in the USA rose in the 1970s, and, after the results of NCDS were published, reached 245.6/1000 patients at risk in 1988 [21]. Of particular interest is the fact

**Table 2. Traumatic interocular tests in dialysis**

| Indication | Intervention | Number of successful cases | Year | Reference |
|---|---|---|---|---|
| Acute renal failure | Hemodialysis | 1 of 16 | 1945 | [46] |
| Chronic renal failure | Hemodialysis | 2 of 2 | 1960 | [47] |
| Blood access | Arteriovenous fistula | 12 of 14 | 1966 | [48] |
| Peritoneal access | Cuffed catheter | 6 of 6 | 1968 | [49] |
| Chronic renal failure | Continuous ambulatory peritoneal dialysis | 9 of 9 | 1978 | [50] |

that it was observed 20 years after NCDS data were published: 'It is difficult to conclude from the NCDS that session length is not meaningful; … one might argue that NCDS session length P = 0.06 was the most significant (important) "non-significant" (statistically) effect in the history of dialysis research' [22]. Contrary to the observational studies leading to improved dialysis results, this randomized controlled study may have adversely impacted survival on HD in the USA.

Data from the European and Japanese registries indicated much better survival in association with longer weekly time on dialysis in Europe and Japan. During a conference in Dallas in 1989 [23], several presentations indicated that the difference in mortality was not related to the patient mix or other factors but shorter dialysis time in the USA [24]. Shortly thereafter, Held *et al.* [25] reported that the prescribed HD dose in Europe was substantially higher than in the USA. After these results of observational studies, a goal $Kt/V_{urea}$ index of 1.2–1.3 was suggested.

In 1995, the National Kidney Foundation-Dialysis Outcomes Quality Initiative (NKF-DOQI) developed guidelines for improved patient outcomes and survival by providing recommendations for optimal clinical practices [26]. Adequacy of dialysis was established as a sp$Kt/V_{urea}$ of 1.2 (single-pool) for all HD patients dialyzed thrice weekly, regardless of age and comorbid conditions [26]. To guarantee the minimum dose of 1.2, a value of 1.3 was recommended as a prescribed minimum [26]. In 2000, the NKF K/DOQI update [27] reiterated an adequacy sp$Kt/V_{urea}$ target of 1.3 (single-pool) and e$Kt/V_{urea}$ of 1.05 (double-pool or equilibrated).

All these reports stimulated a discussion on whether the recommended $Kt/V$ should be increased over 1.3. This dilemma led to the second large NIH-sponsored randomized controlled trial on HD outcomes, the HD (HEMO) study. The results were published in 2002 [28]. Out of 2677 screened patients, 1846 were randomized between March 1995 and October 2000. Patients who could not achieve an e$Kt/V_{urea}$ of more than 1.3 within 4.5 h were excluded. The time of dialysis in the standard–dose group was 190 ± 23 min and in the high-dose group 219 ± 23 min. Dialyzer blood flow in the low-dose group 311 ± 51 mL/min and in the high-dose group was 375 ± 32 mL/min. Single-pool and equilibrated $Kt/V_{urea}$ were 1.32 ± 0.09 and 1.16 ± 0.08, respectively, in the standard-dose group. In the high-dose group, these values were 1.71 ± 0.11 and 1.53 ± 0.09, respectively. The conclusion of this study was that 'patients undergoing hemodialysis thrice weekly appear to have no major benefit from a dialysis dose higher than that recommended by current US guidelines.' It is worth noting that the higher dialysis dose was achieved by a 20% increase in blood flow rate and 15% increase in dialysis time. Taking these facts into account, the conclusion of the HEMO study should add 'if the higher dose is achieved mainly by increased dialyzer blood flow.' Contrary to the numerous observational studies showing that longer dialysis time is beneficial for the dialysis outcomes [29–31], there is no study showing that the higher dialyzer blood flow is not detrimental to the outcome on dialysis. As a matter of fact, at least one study indicated that lower pre-pump negative pressures, related mainly to higher dialyzer blood flow, lead to increased hemolysis [32]. Increased hemolysis during dialysis cannot be considered harmless. Thus, the 20% increase in dialyzer blood flow may have negated the benefit of the 15% increase in dialysis time.

The results from the Japanese dialysis registry [33] including over 50 000 patients showed a time of dialysis of below 5 h as an important predictor of death. Therefore, combining '*K*' and '*t*' for the measurement of dialysis dose is inappropriate. In our opinion, the HEMO study committed a Type III statistical error—asking the wrong question and achieving the correct answer: if higher $Kt/V$ is achieved mainly by increased blood flow ($K$), the beneficial effect of such an increase in the $Kt/V$ may not be realized.

The major problem with $Kt/V_{urea}$ as the index of dialysis quality is the fact that urea is relatively non-toxic and one of the molecules most rapidly transported between body fluid compartments [34]. Urea is a small (60 D), uncharged molecule and transported between fluid compartments through aquaporins. Charged molecules such as sodium (positively charged), phosphate (negatively charged) and guanidino acetic acid (negatively charged) are transported at a much lower rate and some require help with their transport. Bigger uncharged molecules such as β-microglobulin, are also transported slowly between compartments. The slow transport of sodium is particularly important for blood pressure control and dialysis hypotensive episodes during short sessions with a high ultrafiltration rate (UFR). In 2006, a DOPPS observational study [35] determined that (i) the duration of an HD session is independently associated with a lower mortality risk after adjustment for case mix, dialysis dose ($Kt/V$), body size measures and indicators of non-adherence; (ii) delivering a high $Kt/V$ over longer treatment time (TT), up to 270 min, is of greater value than delivering the same $Kt/V$ over shorter TT and (iii) UFR >10 mL/h/kg body weight is independently associated with higher risk of both intradialytic hypotension and mortality. Interestingly, the reanalysis of the HEMO study regarding UFRs and mortality concluded that the UFR over 10 mL/h/kg body weight was associated with increased cardiovascular and overall mortality [36]. More observational papers have been recently published indicating that shorter dialysis time is associated with increased mortality [37].

What about the frequency of dialysis? In the late 1960s and the early 1970s, there were studies, either with historical controls or based on the 'before and after method' showing that more frequent HDs were superior to less frequent ones [17, 38, 39]. The number of centers performing frequent (quotidian) HDs was steadily increasing as it became obvious to the physicians observing more frequently dialyzed patients that this method is better. In 1998, Kjellstrand [40] reviewed the reports on more frequent HD from Brazil, Belgium, Canada, Germany, Finland, Poland, USA and many Italian centers. The outcomes were all very similar and none of these reports showed worse results with more frequent dialysis. The conclusions were based on before-and-after designs so the mortality could not be compared.

There was no question about the superiority of frequent (5–7 times weekly) HD to routine (thrice weekly) dialysis. Unfortunately, more frequent dialysis was more expensive so

those involved in daily (frequent) dialysis wanted increased reimbursement from Medicare. Prior to making a decision on this issue, NIH summoned a Task Force on Daily Dialysis in Washington, DC, on 11April 2001. In spite of the arguments of those who personally observed excellent clinical and laboratory results of more frequent dialysis, the conference decided that observational studies are unreliable and RCTs would be needed to justify higher reimbursement. According to physicians practicing daily dialysis, the probability that numerous observational studies showing improved results with more frequent HD were wrong was close to zero (if, compared with RCTs). Originally, it was assumed that mortality would be considered a primary outcome in such a trial. The reality turned out to be different. Within a few years, it was obvious that recruitment would be a problem. In 2007 Suri *et al.* [41] admitted that problems with recruitment of subjects forced them to abandon comparison of mortality. The problem with recruitment of volunteers for such a study should not be surprising as it was impossible to compare results of peritoneal dialysis and HD in a RCT [42].

Anyway, the prospective randomized study comparing more frequent dialysis with conventional dialysis was started by the Frequent Hemodialysis Network group, lasted almost 10 years, cost millions of dollars and the results were published in December 2010 [43]. Only 245 patients could be randomized and randomization did not include the average patient population as mortality in the control (conventional dialysis) group was only 7.5%, whereas mortality in the US population was over 18.5%. The patients already on more frequent dialysis were excluded, which created another selection bias. The patients on more frequent HD were dialyzed in centers on regular machines instead of machines suitable for more frequent HD at home. At least this RCT did not commit a Type II or a III statistical error; however, it committed a Type IV error, 'solving the problem too late.'

The fourth RCT in HD, fraught with the usual problems faced when such trials are undertaken (inadequate enrollment) resulting in misleading results, was recently published [44]. Only 87 patients could be randomized for this study. Contrary to the study of frequent short dialysis, where patients with substantial residual renal function were excluded from randomization, 57.2% of patients in the control group had a urine output of over 500 mL/day (including 19.1% who had urine output over 1000 mL/day). Patients with a urine output of over 1000 mL/day do not require high-dose dialysis. In the past, such patients were not dialyzed, until their urine output dropped to <1000 mL/day (17). Further indication of an inappropriate selection of patients for randomization is the fact that mortality in the conventional arm was 1/42 = 2.38%, which is at least seven times lower than in the general population of HD patients in the USA. Many patients in the control group performed more frequent dialysis sessions and those in the more frequent group performed less frequent dialyses. The ultimate conclusion that was reached was that the frequent nocturnal dialysis study group had improvement in 'control of hyperphosphatemia and hypertension but no benefit among other main secondary outcomes.' All in all, this study committed a Type II statistical error because of evidently small number and inappropriate selection of subjects. On the basis of this study, it absolutely cannot be accepted that frequent nocturnal HD is not better than conventional thrice-weekly HD. Interestingly, Kidney International, where this study was published, highlights the rather misperceived notion that frequent nocturnal HD is not better all too glaringly on its cover: 'No benefit from frequent nocturnal HD.'

## CONCLUSIONS

All progress in dialysis methods was made in research presented in case reports, case–control studies and other observational studies. Comparison of the value of peritoneal dialysis and HD in RCTs was not completed because of recruitment problems. Four RCTs in HD did not provide any useful data. The worst example was the first (NCDS), which committed a Type II statistical error rejecting the time of dialysis as an important factor determining the quality of dialysis. This study also provided the basis for the establishment of the *Kt/V* index as a measure of dialysis adequacy. This index was accepted by the HD community, as having been established in a sacrosanct RCT, led to short dialysis, and possibly higher mortality in the USA. The second trial (the HEMO study) committed a Type III statistical error by asking the wrong question, did not bring any valuable results, but did not lead to deterioration of dialysis outcomes in the USA. The third [43] did not bring forth any valuable results, but at least confirmed what was already known. The fourth [44] did not bring any positive results because of tremendous recruitment problems leading to inappropriate selection of patients

As observed in a recent editorial [45], 'observational studies allow clinical research to represent the full breadth of treated patients and offer tremendous power … .' The key is to realize that the 'design and conduct of RCTs need to be more inclusive and efficient' [45]. In the field of dialysis, this may be especially difficult to achieve.

## CONFLICT OF INTEREST STATEMENT

None declared.

## REFERENCES


1. Feinstein AR. Statistical significance versus clinical importance. *Qual Life Cardiovasc Care. Autumn* 1988; 99–102
2. Rawlins MD. De testimonio. On the Evidence for Decisions about the Use of therapeutic Interventions. The Harveian Oration. The Royal College of Physicians. Salisbury, Wiltshire: Sarum ColourView Ltd, 2008
3. Feinstein AR. Commentary: two centuries of conflict-collaboration between medicine and mathematics. J Clin Epidemiol 1996; 49: 1339–1443

4. Wright DB. Understanding Statistics: an Introduction for the Social Sciences. London, Thousand Oaks, New Delhi: Sage Publications, 1997, p. 42
5. Fisher RA. Statistical Methods for Research Workers. Edinburgh: Oliver and Boyd, 1925
6. Neyman J, Pearson ES. 'On the Problem of the Most Efficient Tests of Statistical Hypotheses'. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, Vol. 231, 1933, pp. 289–337
7. Kimball AW. Errors of the Third Kind in Statistical Consulting. J Am Stat Assoc 1957; 52: 133–142
8. Raiffa H. Decision Analysis: Introductory Lectures on Choices under Uncertainty. Reading, MA: Addison-Wesley, 1968. p. 264
9. Fisher RA, Mackenzie WA. Studies in crop variation. II. The manurial response of different potato varieties. J Agric Sci 1923; 13: 311–320
10. A Medical Research Council Investigation. Streptomycin treatment of pulmonary tuberculosis. Br Med J 1948; 2: 769–782
11. Hill AB. Heberden Oration 1965: reflections on the controlled trial. Ann Rheum Dis 1965; 25: 107–113
12. Doll R. On the etiology of lung cancer. J Natl Cancer Inst 1950; 11: 638–640
13. Doll R, Hill AB. Smoking and carcinoma of the lung; preliminary report. Br Med J 1950; 2: 739–748
14. Fisher RA. 'Smoking: The Cancer Controversy'. Edinburgh: Oliver and Boyd, 1959, pp. 45–47
15. Sribner BH. Personalized history of chronic hemodialysis. Am J Kidney Dis 1990; 16: 511–519
16. Curtis FK, Cole JJ, Fellows BJ *et al.* Hemodialysis in the home. Trans Am Soc Artif Intern Organs 1965; 11: 7–10
17. Twardowski Z. Effect of long-term increase in the frequency and/or prolongation of dialysis duration on certain clinical manifestations and results of laboratory investigations in patients with chronic renal failure. Acta Med Pol 1975; 16: 31–44 Republished in Hemodial Int. 2004 Jan 1; 8(1):30–38
18. Lowrie EG, Laird NM, Parker TF *et al.* Effect of the dialysis prescription of patient morbidity: report from the National Cooperative Dialysis Study. N Engl J Med 1981; 305: 1176–1181
19. Parker TF, Laird NM, Lowrie EG. Comparison of the study groups in the National Cooperative Dialysis Study and a description of morbidity, mortality, and patient withdrawal. Kidney Int Suppl 1983 Apr; (13); S42–S49
20. Gotch FA, Sargent JA. A mechanistic analysis of the National Cooperative Dialysis Study (NCDS). Kidney Int 1985; 28: 526–534
21. US Renal Data System. USRDS 2011 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States. Bethesda, MD: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, 2011
22. Chertow GM, Kurella M, Lowrie EG. The tortoise and hare on hemodialysis: does slow and steady win the race? Kidney Int 2006; 70: 24–25
23. The Morbidity, Mortality and Prescription of Dialysis Symposium. Proceedings Dallas, Texas, September 15–17, 1989. Am J Kidney Dis 1990; 15: 375–515
24. Held PJ, Brunner F, Odaka M *et al.* Five-year survival for end-stage renal disease patients in the United States, Europe, and Japan, 1982 to 1987. Am J Kidney Dis 1990; 15: 451–457
25. Held PJ, Blagg CR, Liska DW *et al.* The dose of hemodialysis according to dialysis prescription in Europe and the United States. Kidney Int Suppl 1992; 38: S16–S21
26. NKF-DOQI Clinical Practice Guidelines. II Hemodialysis Dose. Am J Kidney Dis 1997; 30(3 suppl 2): S32–S37
27. National Kidney Foundation. K/DOQI Clinical Practice Guidelines for Hemodialysis Adequacy, 2000. Am J Kidney Dis 2001; 37(1 suppl 1): S7–S64
28. Eknoyan G, Beck GJ, Cheung AK *et al.* Hemodialysis (HEMO) Study Group. Effect of dialysis dose and membrane flux in maintenance hemodialysis. N Engl J Med 2002; 347: 2010–2019
29. Valderrábano F. Weekly duration of dialysis treatment—does it matter for survival? Nephrol Dial Transplant 1996; 11: 569–572
30. Locatelli F, Manzoni C. Duration of dialysis session—was Hegel right? Nephrol Dial Transplant 1999; 14: 560–563
31. Lameire N. Management of the hemodialysis patient: a European perspective. Blood Purif 2002; 20: 93–102
32. Twardowski ZJ, Haynie JD, Moore HM. Blood flow, negative pressure, and hemolysis during hemodialysis. Home Hemodial Int 1999; 3: 45–50
33. Shinzato T, Nakai S, Akiba T *et al.* Survival in long-term haemodialysis patients: results from the annual survey of the Japanese Society of Dialysis Therapy. Nephrol Dial Transplant 1997; 12: 884–888
34. Eloot S, Torremans A, De Smet R *et al.* Kinetic behavior of urea is different from that of other water-soluble compounds: the case of guanidino compounds. Kidney Int 2005; 67: 1566–1575
35. Saran R, Bragg-Gresham JL, Levin NW *et al.* Longer treatment time and slower ultrafiltration in hemodialysis: associations with reduced mortality in the DOPPS. Kidney Int 2006; 69: 1222–1228
36. Flythe JE, Kimmel SE, Brunelli SM. Rapid fluid removal during dialysis is associated with cardiovascular morbidity and mortality. Kidney Int 2011; 79: 250–257
37. Lacson E, Brunelli SM. Hemodialysis treatment time: a fresh perspective. Clin J Am Soc Nephrol 2011; 6: 2522–2530
38. DePalma JR, Pecker EA, Maxwell NH. A new automatic coil dialyzer system for 'daily' dialysis. Proc Eur Dial Transplant Assoc 1969; 6: 26–34
39. Bonomini V, Mioli V, Albertazzi A *et al.* Daily dialysis programme: indications and results. Proc Eur Dial Transplant Assoc 1972; 9: 44–52
40. Kjellstrand CM. A brief history of daily hemodialysis. Home Hemodial Int 1998; 2: 8–11
41. Suri RS, Garg AX, Chertow GM *et al.* Frequent Hemodialysis Network Trial Group. Frequent Hemodialysis Network (FHN) randomized trials: study design. Kidney Int 2007; 71: 349–359 Epub 2006 Dec 13
42. Korevaar JC, Feith GW, Dekker FW *et al.* NECOSAD Study Group. Effect of starting with hemodialysis compared with peritoneal dialysis in patients new on dialysis treatment: a randomized controlled trial. Kidney Int 2003; 64: 2222–2228
43. FHN Trial GroupChertow GM, Levin NW, Beck GJ *et al.* In-center hemodialysis six times per week versus three times per week. N Eng J Med 2010; 363: 287–300

44. Rocco MV, Lockridge RS, Jr, Beck GJ *et al.* the Frequent Hemodialysis Network (FHN) Trial Group. The effects of frequent nocturnal home hemodialysis: the Frequent Hemodialysis Network Nocturnal Trial. Kidney Int 2011; 80: 1080–1091
45. Mauri L. Why we still need randomized trials to compare effectiveness. N Engl J Med 2012; 366: 1538–1540
46. Kolff WJ. First Clinical Experience with the Artificial Kidney. Ann Intern Med 1965; 62: 608–619
47. Scribner BH, Buri R, Caner JEZ *et al.* The treatment of chronic uremia by means of intermittent hemodialysis: a preliminary report. Trans Am Soc Artif Intern Organs 1960; 6: 114–122 Republished in J Am Soc Nephrol 1998; 9: 719–726
48. Brescia MJ, Cimino JE, Appel K *et al.* Chronic hemodialysis using venipuncture and a surgically created arteriovenous fistula. N Engl J Med 1966; 275: 1089–1092 Republished in J Am Soc Nephrol 1999 Jan; 10(1): 193–199
49. Tenckhoff H, Schechter H. A bacteriologically sate peritoneal access device. Tans Am Soc Artif Intern Organs 1968; XIV: 181–186
50. Popovich RP, Moncrief JW, Nolph KD *et al.* Continuous ambulatory peritoneal dialysis. Ann Intern Med 1978; 88: 449–456 Republished in JASN 1999; 10 (4): 901–910

## Opponent's comments

It is gratifying that Twardowski and Misra accept the primacy of experimentation to test the null hypothesis. It is also important that they accept that they have "absolutely nothing against properly conducted RCTs".

However, they provide an emotional argument that "all randomized controlled trials (RCTs) in dialysis failed owing to either inadequate planning or due to misplaced belief that they would be easy to conduct and complete." Of course, this is untrue! The trials may not have supported the well-intentioned belief of Twardowski and Misra that more dialysis is better but, as I have pointed out in my rebuttal, these studies provided important insights.

Twardowski and Misra take issue with me in my support of the intention to treat principle as a central feature of a properly conducted randomized trial. Of course, I am not alone in this since the ITT principle is a core feature of virtually every properly conducted RCT.

Twardowski and Misra state: "Intention to treat analysis is the major weakness of RCTs. How can one deny such a bias if a patient allocated to interventional group undergoes a treatment meant for the control group but such patient's data are included in the interventional group?" Twardowski and Misra remind me of the old adage "throwing the baby out with the bathwater". Twardowski and Misra are correct in pointing out that cross-overs from one treatment assignment to the other is a limitation to a properly conduced RCT. It was in FHN but no trial is perfect.

Likewise, selection bias, lack of power, an imbalance between the two or more randomized arms of a study, or an excessive drop-out rate represent examples where an RCT may have limitations. Perfection should not be the enemy of good. Applying the ITT principle in a flawed study does increase the possibility of a type II error but this needs to be balanced against it's main advantage in minimizing the possibility of a type I error (1,2). The cautiousness inherent in using the ITT principle ultimately allows for the greatest generalizability. Indeed, as others have stated (3), "ITT analysis avoids overoptimistic estimates of the efficacy of an intervention resulting from the removal of non-compliers by accepting that noncompliance and protocol deviations are likely to occur in actual clinical practice."

Twardowski and Misra have made many important contributions to our understanding of dialysis dose and modality; however, respectfully, their current argument rests on rather thin ice.

Ajay K. Singh

## References

1. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: The intention to treat principle and excluding patients from analysis. BMJ. 2002;325:652–4.
2. Hollis S, Campbell F. What is meant by intention to treat analysis. Survey of published randomised controlled trials? BMJ. 1999;319:670–4.
3. Gupta SK Intention-to-treat concept: A review Perspect Clin Res. 2011 Jul-Sep; 2 (3): 109–112.

*Received for publication: 27.3.2012; Accepted in revised form: 11.5.2012*