Tacrolimus (FK506) and the Pharmaceutical/Academic/Regulatory Gauntlet

Thomas E. Starzl, MD, PhD, Satoru Todo, MD, Anthony J. Demetris, MD, and John J. Fung, MD, PhD

• The pivotal issue of transplant rejection diagnosis and management is design, conduct, and analysis of clinical trials. The historical experience with clinical trials of major immunosuppressive drugs (cyclosporine and especially tacrolimus) is examined in this article. Cyclosporine was a turning point in transplantation, providing an extraordinary improvement over previous therapies. Additionally, early investigational experience with tacrolimus was shown to be important in rescue from cyclosporine failure. Experience with tacrolimus in liver recipients for primary therapy led to understanding that the side effect profile was similar to cyclosporine and that the important side effects of tacrolimus (toxicity and diabetes) could be lessened by altering the drug dose. Early dosing regimens were determined by attempts to balance the toxicities (representing a dose ceiling) against rejection (for minimum dosing). Drug levels became understandable and trough levels could be used to guide therapy. However, when the multicenter liver trial was implemented, high starting doses were included in the protocol design, ignoring information obtained with drug level monitoring. Disregard for this information led to a distortion of the potential efficacy, and randomized trials were required by the FDA. The fact that multicenter trials examined thoroughly. When tacrolimus was proposed for use in extrahepatic organ transplantation, again the Pittsburgh experience, as well as experience from other single centers, was determined inadequate evidence of efficacy, and randomized trials were required by the FDA. The fact that multicenter trials in transplantation have historically been poorly designed or analyzed weighed against the dramatic improvements shown from historically controlled studies or single-center trials should lead to question of the regulatory requirement for multicenter randomized trials for all organ types.

© 1998 by the National Kidney Foundation, Inc.

INDEX WORDS: Kidney; liver; transplantation; tacrolimus; cyclosporine; regulatory; clinical trials.

MUCH OF THE discussion at this consensus symposium on immunosuppressive drug testing has concerned the clinical and/or histopathological diagnosis of rejection. As important as this may be, it is not the core issue that we should be examining. Most errors in the evaluation of new immunosuppressants can be traced to the design, conduct, and analysis of clinical trials. In turn, the scientific considerations that go into what is indisputably a human experiment are inseparable from the ethical context into which they are placed.1-3 Questions about the probity of trials with experimental agents to control rejection were not commonly asked until 15 years ago for the simple reason that the historical results were so unsatisfactory. With the advent of cyclosporine, the permissive attitude about deviations from preexisting routine hardened to reflex skepticism and resistance.

CYCLOSPORINE BENCHMARK

When cyclosporine was first used clinically as monotherapy or in combination with myelotoxic agents,4 multiple serious side effects were observed. These were brought into an acceptable range by combining cyclosporine with dose-manageable prednisone5 and later with the addition of third, fourth, and even fifth agents. These were more potent versions of the previous best azathioprine/prednisone6 or azathioprine/antilymphocyte globulin (ALG)/prednisone cocktails7 that had been in common use for almost two decades. The cyclosporine-based recipes that were developed with kidney transplantation improved the prognosis of all organ recipients and elevated liver transplantation overnight to practical and widespread use.8

From the Pittsburgh Transplant Institute and the Departments of Surgery and Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA.
Aided by Project Grant No. DK 29961 from the National Institutes of Health, Bethesda, MD.
Address reprint requests to Thomas E. Starzl, MD, PhD, Department of Surgery, 3601 Fifth Ave, SC Falk Clinic, University of Pittsburgh, Pittsburgh, PA 15213. E-mail: mangan@med.pitt.edu
© 1998 by the National Kidney Foundation, Inc.
0272-6386/98/3106-0102$3.00/

TACROLIMUS

Rescue Phenomenon

Cyclosporine remained the unchallenged baseline immunosuppressant for all varieties of transplantation until it was shown in 1989 that intractably rejecting liver allografts could be regularly rescued by replacing cyclosporine with tacrolimus. Since this original report, numerous confirmatory descriptions have been published. It also was promptly recognized that tacrolimus could salvage rejecting hearts, kidneys, and other organs as well. In virtually all such cases, a switch back to cyclosporine was never made.

Pilot Primary Treatment Experience

By early 1990, more than 150 liver, kidney, heart, and heart-lung recipients had been treated in Pittsburgh with tacrolimus rather than cyclosporine from the time of transplantation. It had been learned that the three major side effects of the drug (nephrotoxicity, neurotoxicity, and diabetogenicity) were comparable to cyclosporine. Hypertension and hyperlipidemia were less than in historical cyclosporine controls, and the cosmetic effects of cyclosporine (hirsutism, gingival hyperplasia, and facial brutalization) had not been seen. Although it was recognized that these first patients had been started on too much tacrolimus, complications of overdosage had been minimized by using the characteristic side effects cited above to determine dose ceilings from the first day of treatment onward. Rejection established the dose floor. From these observations, the meaning of drug plasma level concentrations was quickly deduced. This allowed drug trough level monitoring to be exploited in subsequent cases. This sequence of development was no different than that followed a decade earlier with cyclosporine.

In addition, the availability of an on-site plasma assay laboratory in Pittsburgh (the only one outside of Japan) allowed the pharmacokinetics of tacrolimus in humans to be delineated by early 1990. Relative to cyclosporine, the new drug's absorption was disturbed very little by the absence of bile or by intestinal disorders. However, its rapid elimination was more dependent on good liver function. As a consequence of both factors, recipients of poorly functioning hepatic grafts had been observed to have plasma trough concentrations nearly 50 times those considered optimal. Even when the hepatic grafts functioned well, liver recipients required on average only slightly more than half the dose of kidney recipients to achieve equivalent trough plasma levels.

MULTICENTER LIVER TRANSPLANT TRIALS

Although all of the foregoing information was widely known at the time the multicenter liver trials began in late August and late September 1990, group decisions were made on both sides of the Atlantic to use the high starting doses of tacrolimus that already had been abandoned in the pilot center. Within a few weeks after the trials began, a worldwide epidemic of toxicity reports resulted. Formal dose revisions were not made until 30% and 18% of the European and American tacrolimus case enrollment, respectively, had occurred. The gap between the multicenter study starting doses and those in concurrent use in Pittsburgh never closed, even by the end of these trials (Fig 1).

The multicenter investigators were themselves the architects of the treatment protocols after consultation with the government regulatory agencies and officials of the sponsoring drug company (Fujisawa Pharmaceutical Co, Ltd, Osaka, Japan). No familiarization (pilot) cases were allowed using the drug from the time of transplantation. Despite this handicap, the systematic error in dosing, and the disadvantage of not having on-site assay for drug monitoring of tacrolimus...
limus, the investigators who had vast cumulative experience with cyclosporine-based immunosuppression salvaged the trials. This accomplishment showed how quickly a talented group of clinician-investigators in 20 different centers (12 American, 8 European) could surmount a learning curve for an experimental drug and introduce flexibility into excessively rigid management protocols that cannot be applied in exactly the same way to any two recipients. It also demonstrated that the drug was user-friendly. The down side, however, was that a distorted picture of the new drug’s potential value emerged, in contrast to that clearly delineated in the earlier Pittsburgh experience.

DESIGN AND OUTCOME OF LIVER TRANSPLANTATION TRIALS

Historically Controlled Pittsburgh Series

In the 32 years of this program (which had begun at the University of Colorado), only two major improvements in patient survival have occurred from the level established with the cocktail of azathioprine, prednisone, and ALG. The first came in 1980 with the advent of cyclosporine, and the second followed the introduction of tacrolimus. A small increment temporarily associated with the availability of the University of Wisconsin preservation solution in 1987 was dependent on the delayed rescue with tacrolimus of patients ailing on cyclosporine from the immediately preceding era.

Improved survival of grafts was almost of the same magnitude as the jump that had followed the change from azathioprine to cyclosporine (Fig 2). With primary tacrolimus treatment, graft loss from refractory rejection occurred in only 1% to 2% of cases. Of the 1,391 liver recipients entered, only 35 (2.5%) crossed over from tacrolimus to cyclosporine; of these, 15 changed back when rejection supervened.

Pittsburgh Randomized Trial

In this single-center trial, tacrolimus and cyclosporine were compared head to head. All treatment variables other than discretionary use of the competing drugs were equal at the outset including a daily dose of 20 mg prednisone (Fig 3). Thus, the occurrence of rejection and the need to treat it with additional prednisone or other adjuvant therapy directly reflected the efficacy of the competing drugs. Treatment failure was precisely defined by the inability to control biopsy-confirmed rejection with an orderly sequence of escalating secondary intervention. If these maneuvers were unsuccessful, switch to the other baseline drug was permitted. The side effects and other complications were automatically recorded. Crossover in either direction was permitted but discouraged.

The trial (February 1990–November 1991) was characterized by the rapid movement of patients off the cyclosporine arm. Eventually, 47 of the 75 patients randomized to cyclosporine switched, triggered by treatment failure under steroid-sparing conditions of the protocol. With the rescue capabilities of tacrolimus, 1-year patient survival (with intent-to-treat analysis) was 94% tacrolimus versus 89% cyclosporine with 1-year graft survival of 90% versus 80% (not included).
significant). At the end of 4 years' follow-up, patient survival was 85% in both arms, and graft survival was 80% tacrolimus versus 77% cyclosporine (not significant). Because the near parity of results in the two arms was so dependent on the rescue qualities of tacrolimus, the use of the intent-to-treat analysis has been questioned. However, there was a significantly lower incidence of rejection with tacrolimus, a reduced need for steroids, and several other quality-of-life advantages.

The study was terminated in 1991 at the recommendation of a multi-institutional Patients' Rights Committee. A full accounting of the trial is provided elsewhere.

**Multicenter Trials**

In addition to the excessive tacrolimus dosing (see earlier), the unbalanced use of secondary immunosuppressants (Fig 4) was scientifically controversial. The cyclosporine arm was uploaded with twice the induction doses of prednisone in all 12 American centers, a third drug (azathioprine) in 10 centers, and a fourth agent (polyclonal ALG) in one. The eight European protocols were similar. On both sides of the Atlantic, cyclosporine dose selection for induction and subsequent adjustment were at the physicians' discretion. In contrast, the high starting doses of tacrolimus were obligatory, and dose adjustments were hampered by delay in drug-monitoring results caused by shipping samples to reference laboratories in distant cities. Drug crossover in either direction was permitted for intractable rejection but only from tacrolimus to cyclosporine for the indication of side effects.

**European.** The trial involved 545 patients. A 5% better patient survival was recorded on the tacrolimus arm (46 v 61 deaths) and a 5% higher graft survival. The survival advantage was not statistically significant, but the authors noted that approximately 10% of the surviving grafts credited to cyclosporine had been rescued with tacrolimus. The distorting roles of tacrolimus overdosage and a high rate of toxicity were partially clarified by separate analyses of the early (high dose) and late (reduced dose) phases of the trial. The statistical analysis, based on the intent-to-treat approach, showed significantly greater freedom from acute rejection, intractable acute rejection, and chronic rejection.

**American.** Of the 529 enrolled patients, only 65% completed the first year of this study (180 tacrolimus, 164 cyclosporine). Although the published report was claimed to be by intent-to-treat analysis, the only analyses actually done by intent-to-treat were patient survival and graft survival, which were not significantly different in the two arms. The rescue role of tacrolimus in reducing the overall incidence of retransplantation was obliquely acknowledged in the discussion: "The low number of second transplantations for refractory rejection may have been due, in part, to the effectiveness of tacrolimus in treating patients in the cyclosporine group who had refractory rejection." In fact, grafts rescued by tacrolimus accounted for 20 of the 210 surviving grafts (9.5%) credited by intent-to-treat analysis to the cyclosporine arm at the end of the year.

Extensive case censoring, misuse of the Kaplan-Meier method, and violations of the intent-to-treat principle led to a reanalysis (which has been published elsewhere) of the original database. With reanalysis, freedom from rejection as a single endpoint, as well as combined freedom from all of the secondary endpoints was greater. It was noteworthy that drug toxicity had an almost immeasurable effect on the tacrolimus superiority, correcting the impression left by the original report that the greater tacrolimus efficacy was balanced out by increased toxicity.

The most clinically relevant result of the reanalysis was that after 1 year of follow-up, 98% of the tacrolimus-randomized patients had freedom from the diagnosis of refractory rejection versus only 87% in the competing arm. In addition, the composite freedom at 1 year from the
three factors that haunt transplant recipients (refractory rejection, retransplantation, and death) was 80% tacrolimus versus 70% cyclosporine.30

**KIDNEY AND OTHER ORGAN TRANSPLANTATION**

Historically, kidney transplantation was the whole-organ procedure with which new immunosuppressive drugs and regimens were evaluated.4,6,7,31-33 This precedent was broken with the large-scale randomized liver transplant trials. The decision to go forward with the multicenter trials was made despite prima facie evidence of tacrolimus superiority from the increasingly well-documented rescue of liver allografts (see earlier) and shortly thereafter of intractably rejecting kidneys44 and hearts.35 Remarkably, this experience that was promptly confirmed in multiple other centers was dismissed as a basis for regulatory approval because of the difficulty (and ethical unacceptability) of obtaining “controls.”

Meanwhile, recipients of extrahepatic organs treated with tacrolimus in Pittsburgh (and soon elsewhere) from the time of transplantation were reported to enjoy the same advantages with this drug as the early liver recipients.12,35,36 A randomized renal transplantation trial of tacrolimus versus cyclosporine was carried out in Pittsburgh in 1990 and 199137 but was suspended less than a year after its inception at the recommendation of the same Patients’ Rights Committee that had taken similar action with the liver trial (see earlier).

Although the patient and graft survival in the single-center randomized kidney transplant trial was equal on the two treatment arms, the parity by intent-to-treat analysis was dependent on the salvage of intractably rejecting crossover patients rescued with tacrolimus as well as on higher average steroid doses in patients kept on cyclosporine. The same results were obtained later in single-center trials of lung18 and heart transplantation.39 Surprisingly, with this wealth of information, a second regulatory precedent was broken by an FDA decision that randomized trials would be mandatory for every kind of transplantation, organ by organ, rather than considering the lessons from the liver and other transplant trials to be generic.

**DISCUSSION AND CONCLUSION**

Disquieting scientific and social issues were exposed by the events during the 6 years following the placement of tacrolimus on the FDA “fast track” in November 1989. The issue of the new drug’s unusual rescue capability was never in doubt, and the question of its superiority as the primary baseline agent was scarcely less clear by the time the multicenter liver trials were started. It is not unreasonable, therefore, to ask why the multicenter liver trials were performed and why such trials with other organs are still ongoing.

Although a properly designed and executed randomized controlled trial (RCT) is the best way to determine the effectiveness of competing patient therapies, such studies have been under-represented in the surgical literature.40,41 A partial explanation has been provided by Solomon and McLeod42 who found that even in an ideal clinical research setting only 40% of questions involving a surgical procedure could have been answered by a correctly formulated RCT. Although they generally favor such studies whenever feasible, the authors stated that “… if a new treatment is shown to result in a dramatic improvement in outcome in uncontrolled, immediate historic controlled trials or non-RCT, a RCT may be unnecessary or even unethical.”42

Was there such an absence of equipoise at the time of the multicenter liver trials? Almost certainly yes, if the observations from Pittsburgh on rescue as well as baseline therapy with tacrolimus were valid. Independent confirmation was required at other centers. The ostensible options from which the FDA could have selected for this purpose are listed in Table 1. However, the agency has increasingly insisted on multicenter controlled randomized trials as a prerequisite for marketing new drugs, and this was the decision with tacrolimus. It has been repeatedly stated and often persuasively argued that inappropriate and/or ineffective therapy can become institutionalized without such trials on the basis of case series and retrospective studies.

The randomized trial policy has powerful support in university circles, for reasons that go well beyond its intellectual merit. Fiscal, administrative, and professional opportunities are generated within each component of the “regulatory/pharmaceutical/academic complex” that drives...
such trials. The consequent range of possible conflicts of interest has made randomized trials a magnet for criticism. The most damaging potential allegation has been that such studies are frequently performed to obtain answers that are already known1-3; the multicenter tacrolimus/cyclosporine liver and more recent other organ trials are prime examples.30

Reform in the evaluation of immunosuppressive drugs will have to begin with the FDA for two reasons. The sole purpose of this agency is public service, unlike the other participants in the pharmaceutical/academic/regulatory triad. Second, the FDA is self-empowered to select the requisite evidentiary pathways (meaning what controls) that can lead a drug to the market place (Table 1). The FDA does not ostensibly engage in human experimentation and recoils reflexively at the allegation. However, when it determines that a multicenter randomized trial, rather than one of the other available options (such as historical controls), is a condition for sale of a new drug, the agency becomes the de facto instigator of a human experiment and the silent partner of the investigating physicians, Institutional Review Boards, and pharmaceutical companies who must supervise or perform the study. It is not possible to pull this switch and disavow responsibility for what follows.

Can ill-advised or poorly designed randomized trials have an effect opposite to the objectives of improved and less-expensive patient care? The answer involves more than the consequent increase in drug cost that eventually is passed on to the patients and public. Gjertson, Cecka, and Terasaki43 recently reported evidence from 24 American kidney transplant centers with access to tacrolimus that the projected actuarial half-life of cadaver renal allografts was 14 years in recipients treated from the outset with this drug versus 8 years using any previously available immunosuppressant including cyclosporine. The investigators concluded: “Based on this study, FK506 (tacrolimus) appears to be the first therapeutic agent [in the history of the field] to significantly improve long-term kidney graft survival rates.”

If these projections prove to be valid, the cumulative unnecessary expenses assumed by the taxpayer during the 5-year delay in use of tacrolimus for renal transplantation engendered by current regulatory policies will have been almost beyond imagination. During this period, approximately 25,000 primary renal transplantations were reported to the United Network for Organ Sharing Scientific Registry (the data source for the study43), almost all under cyclosporine. Adding a conservative 5 years’ graft function to each successful transplantation (at 1 year, tacrolimus 91% and cyclosporine 87%) and assuming that the cost of returning to dialysis exceeds that of late posttransplant care by $10,000/yr, the potential cost savings that have been lost, even taking into account patient mortality, calculates well in excess of one billion dollars.

NOTE ADDED IN PROOF

March 26, 1998: This article was submitted at the time of the consensus development meeting on May 13, 1995. Except for completion of citations then in press, it has not been changed. It is noteworthy that the experience with tacrolimus accrued since then in many centers has been consistent with the earlier views.

REFERENCES

4. Calne RY, Rolles K. White DJG. Thiru S, Evans DB, McMaster P, Dunn DC, Craddock GN, Henderson RG, Aziz S, Lewis P: Cyclosporin A initially as the only immunosup-

### Table 1. Spectrum of Most to Least "Adequate" Controls

<table>
<thead>
<tr>
<th>Code</th>
<th>Control Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Active treatment concurrent control (randomized, double-blinded)</td>
</tr>
<tr>
<td>2.</td>
<td>Placebo concurrent control (randomized, double-blinded)</td>
</tr>
<tr>
<td>3.</td>
<td>Dose-comparison concurrent control (randomized, 2 doses, often with placebo or active treatment arm)</td>
</tr>
<tr>
<td>4.</td>
<td>Historical control v concurrent treatment (special situations, when outcome is self-evident or in absence of equipoise)</td>
</tr>
<tr>
<td>5.</td>
<td>No treatment concurrent control (randomized, with no treatment arm)</td>
</tr>
</tbody>
</table>

pressants in 34 recipients of cadaveric organs; 32 kidneys, 2 pancreases, and 2 livers. Lancet 2:1033-1036, 1979


42. Solomon MJ, McLeod RS: Should we be performing more randomized controlled trials evaluating surgical operations? Surgery 118:459-467, 1995