

Published in final edited form as:

*Am J Transplant.* 2009 October ; 9(10): 2217–2222. doi:10.1111/j.1600-6143.2009.02802.x.

## Subgroup Analyses in Randomized Controlled Trials: the Need for Risk Stratification in Kidney Transplantation

Martin Wagner, MD<sup>1,2</sup>, Ethan M. Balk, MD, MPH<sup>3</sup>, David M. Kent, MD, MS<sup>4</sup>, Bertram L. Kasiske, MD<sup>5</sup>, and Henrik Ekberg, MD, PhD<sup>6</sup>

<sup>1</sup> Department of Medicine, Division of Nephrology, Tufts Medical Center, Boston, MA, USA

<sup>2</sup> Department of Medicine I, Division of Nephrology, University Hospital Würzburg, Würzburg, Germany

<sup>3</sup> Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

<sup>4</sup> Center for Predictive Medicine Research, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

<sup>5</sup> Department of Medicine, Division of Nephrology, Hennepin County Medical Center, University of Minnesota, Minneapolis, MN, USA

<sup>6</sup> Department of Nephrology and Transplantation, Lund University, University Hospital, Malmö, Sweden

### Abstract

Although randomized controlled trials are the gold standard for establishing causation in clinical research, their aggregated results can be misleading when applied to individual patients. A treatment may be beneficial in some patients, but its harms may outweigh benefits in others. While conventional one-variable-at-a-time subgroup analyses have well-known limitations, multivariable risk-based analyses can help uncover clinically significant heterogeneity in treatment effects that may be otherwise obscured. Trials in kidney transplantation have yielded the finding that a reduction in acute rejection does not translate into a similar benefit in prolonging graft survival and improving graft function. This paradox might be explained by the variation in risk for acute rejection among included kidney transplant recipients varying the likelihood of benefit or harm from intense immunosuppressive regimens. Analyses that stratify patients by their immunological risk may resolve these otherwise puzzling results. Reliable risk models should be developed to investigate benefits and harms in rationally designed risk-based subgroups of patients in existing RCT datasets. These risk strata would need to be validated in future prospective clinical trials examining long term effects on patient and graft survival. This approach may allow better individualized treatment choices for kidney transplant recipients.

---

Corresponding author: Martin Wagner, MD, University Hospital Würzburg, Department of Medicine I, Division of Nephrology, Zentrum Innere Medizin, Oberdürrbacher Strasse 6, 97080 Würzburg, Germany, Phone: +49 (931) 201-0, Fax: +49 (931) 201-639300, wagner\_m@klinik.uni-wuerzburg.de.

**Financial Disclosure:** Dr. Wagner has received travel grants from Hoffman-La Roche, Germany, Dr. Balk and Dr. Kent have nothing to declare.

Dr. Kasiske receives research grants from Bristol-Myers Squibb, Merck-Schering Plough, Wyeth, and serves on an advisory board for Litholink, Inc. He has received consulting or lecture fees from Astellas, Novartis, and Wyeth.

Dr. Ekberg has received consulting or lecture fees from F. Hoffmann-La Roche, Astellas, Novartis, Protein Design Lab, Life Cycle Pharma, Bristol Myers Squibb, Amgen and Hansa Medical.

## Background

Randomized controlled trials (RCT) are the gold standard to measure the effect of an experimental treatment compared to a control treatment. The randomization process provides the basis for comparing interventions independent of other factors.(1) Performing high quality RCTs can be difficult, and the complexity of solid organ transplantation makes RCTs even more challenging. These problems have been addressed recently in a review that discusses important considerations about study populations, eligibility criteria, selection of the control arm, choice of appropriate endpoints, and the analysis plan.(2) However, the authors were critical about subgroup analyses.

While a multitude of immunosuppressive drugs and a variety of regimens have been tested with RCTs in kidney transplantation, evidence remains unclear about which immunosuppressive regimen is best for an individual patient. Kidney transplant recipients represent a very heterogeneous group of patients. It is likely that some patients benefit to a greater extent from a given treatment, whereas others experience only little benefit, or may even be harmed by the treatment. (3) RCTs that report only summary results are unable to address this inter-individual variation in response to therapy.(4) Indeed, it has been shown that summary results may be misleading in that they reflect the treatment effect of a relatively small subgroup of influential patients, rather than reflect the treatment effect of the typical patient enrolled in the trial.(5;6)

In the setting of kidney transplantation there are several trials that show important effects on intermediate outcomes (e.g., acute rejection) that have been strongly and consistently linked with long-term clinical outcome (e.g., graft loss), yet fail to demonstrate any benefits in the clinical outcomes. This article addresses how this paradox may be explained by the presence of heterogeneity of treatment effects and how risk-based analyses could be helpful in systematically exploring and “disaggregating” the overall results within RCTs.

## Treatment Effect Heterogeneity

Results of RCTs represent the average effect of an intervention for an entire group. While the treatment and control groups as a whole are comparable, it is well recognized that the patients typically differ from each other in their baseline risk for achieving the outcome of interest. This results in wide variation of individual patients' likelihood of achieving the outcome and thus responding to therapy.(4;5) For example, a 40 year old patient with kidney failure from polycystic kidney disease who receives a pre-emptive kidney transplant from a 35 year old spouse has a much higher probability of long-term survival and graft survival than a 60 year old patient with type 2 diabetes and cardiovascular disease (CVD), 8 years of hemodialysis, receiving a second deceased donor kidney transplant. In a trial, the effects of the intervention in these two patients would be averaged and the overall trial result would be assumed to apply equally to both. When a trial shows a treatment benefit that is associated with only some harms, there are typically subgroups of patients in whom this treatment is particularly effective (the benefits clearly outweigh the potential harms) and other patients for whom the treatment results in only minor benefit or even net harm.(6)

## Limitations of Subgroup Analyses and How Risk-Based Analyses Can Help

Conventional subgroup analyses, which test the presence versus the absence of specific risk factors (e.g., diabetic vs. nondiabetic recipient, living vs. deceased donor), can be used to investigate varying treatment effects, but there are important limitations to this approach (7;8). These analyses are prone to spurious false positive results, since multiple sequential hypothesis testing presents multiple opportunities for between-group treatment effect differences to occur due to chance. Subgroup analyses are also prone to false negative

results since trials are typically inadequately powered to uncover treatment effect differences, even in the presence of true treatment effect modification. Finally, one-variable-at-a-time subgroup analyses ignore the fact that patients have multiple factors that interact to influence the likelihood of the outcome of interest and the opportunity for treatment benefit. This scenario is surely the case in kidney transplantation.(9) Testing subgroups of patients who differ only by a single variable compares groups of patients who are more similar to each other than different, and thus may fail to show treatment effect differences even when individual patient effects may vary considerably. (6)

Risk-based analyses aim to overcome some of these limitations. A variety of patient characteristics are combined in a risk score, developed by a multivariate prediction tool. Patients can then be stratified according to their baseline risk for achieving the outcome of interest. Multiple studies and actual trial analyses have demonstrated that risk-based analyses are feasible and can identify patients who benefit most, or who are at greatest risk of harm from a treatment.(10–13) To our knowledge, the only therapeutic intervention that the Food and Drug Administration has specifically labeled based on the patient's risk is the use of activated protein C (Xigris®), which was shown to reduce mortality only in high risk sepsis patients, as assessed by APACHE score.(14)

Nevertheless, methodological issues in risk-based analyses still need to be refined, including whether it might ever be appropriate to employ internally developed risk models or whether risk-group specific stopping rules need to be specified. Frequently, baseline risk is not evenly distributed across patients. While “low risk” patients typically comprise the majority of participants in a trial, a relatively small number of “high risk” patients often account for a large number of outcomes, and thus may drive the overall results. Even in trials that aim to investigate relatively homogenous study populations (by excluding so-called “high risk” patients) a considerable degree of baseline risk heterogeneity can be observed (5). Simulations have shown that a trial adequately powered for its main effect will often be reasonably well powered for a risk stratified analysis, even when the discrimination of the risk model shows only moderate performance.(10)

## **Evidence for Risk-Based Treatment Effect Heterogeneity in Kidney Transplantation**

In kidney transplantation, like other medical fields, it is difficult to know whether baseline risk heterogeneity leads to differences in treatment effects because risk-based analyses have not been performed. However, several paradoxes in the results of clinical trials suggest that treatment effect heterogeneity may be important and that risk-based analyses may yield clinically valuable information.

### **The Effect of Acute Rejection on Graft Outcome**

Over the last decade a variety of agents to reduce acute rejection have been tested in kidney transplantation.(15). It is well-established that acute rejection is strongly associated with graft failure (16). When mycophenolate mofetil (MMF) was tested against azathioprine (AZA), it reduced the incidence of biopsy-proven acute rejection by nearly 50% in the first 6 months (17). However, the 3-year results found no significant differences in graft loss or graft function between MMF and AZA. The absence of evidence for differences in graft survival despite a strong decrease in acute rejection appears especially paradoxical since within the trial acute rejection was a strong risk factor for graft loss or death. Patients who developed acute rejection in the first 6 months post-transplantation were much more likely to lose their graft or die within 3 years than those without acute rejection (odds ratio (OR) 3.7, 95% confidence interval (CI) 2.0–7.0).(18)

Similar findings were evident in another trial where either cyclosporine A (CsA) or MMF was discontinued 3 months after kidney transplantation, in a regimen also containing antithymocyte globulin induction and steroids. Although the rate of acute rejection was lower (6% vs. 22%), graft function (glomerular filtration rate [GFR]) was not superior (38 vs. 46 ml/min) in the MMF withdrawal group (i.e., in those who continued CsA) compared to the CsA withdrawal group (with maintenance MMF). Again, the absence of an effect on graft function in the face of a dramatic decrease in acute rejection was surprising considering the strong association between acute rejection and graft function impairment seen within the trial (OR 4.6, 95% CI 1.4–14.5). In this study, triple therapy was required to be reintroduced in patients experiencing acute rejection. Although the incidence of acute rejection was lower in the MMF withdrawal arm, more patients were on triple therapy at the end of the observation period (33% vs. 22%). This was due mainly to biopsy-proven CsA-toxicity, suggesting that the benefits of continued CsA in terms of reduced risk of acute rejection came at the cost of treatment-related nephrotoxicity.(19)

### Theoretical Explanation of the Phenomenon

The absence of evidence for a difference in graft survival can in part be reasonably explained by a lack of statistical power in these trials. A calculation based on the number of patients enrolled in a trial as well as on the treatment and control rates for acute rejection resulted in a predicted difference of approximately 4% in graft loss at 3 years in the MMF trial. A sample size 8- to 10-fold higher than the number actually enrolled would have been needed to assure sufficient power.(20)

However, the observation of no differences in the surrogate endpoint graft function still needs some elaboration. Variability in baseline risk may help elucidate these apparently paradoxical findings. Consider a hypothetical RCT enrolling 400 patients to treatment groups A and B (Table 1). Treatment B results in less acute rejection (18% vs. 30%, 40% relative risk reduction), but mean kidney function is equal in both groups. Since acute rejection is substantially reduced by treatment B and this is a strong risk factor of kidney function, one can hypothesize that treatment B is associated with treatment-related nephrotoxicity that cancels out the treatment benefit from avoiding acute rejection. While the overall net effect of treatment population-wide is null, if the patients have different baseline risks of acute rejection the net effects of the two treatments is likely to be substantially different across risk strata.

Figure 1 illustrates this case. When the population is stratified according to the risk for acute rejection, all subgroups – and thus the entire cohort – experience the same relative risk reduction of 40% from treatment B compared to treatment A (Panel A). Even in this simplified scenario, with a uniform relative risk reduction in the risk of acute rejection, the benefits of treatment B would vary according to the risk of this outcome. Patients at low risk of acute rejection are unlikely to benefit as much as patients at higher risk since their risk of kidney damage from acute rejection is relatively low even in the absence of the stronger therapy. Thus, the improvement in GFR due to reducing acute rejections varies and is dependent on the patients' baseline risk for acute rejection (Panel B). To simplify, let us assume that only treatment B has nephrotoxic effects, which reduce GFR by 10 ml/min in all subgroups (Panel C). Due to the reduction in GFR from treatment B, GFR in the entire cohort is now similar when the component of treatment-related harm is taken into account (Panel D). Patients at low risk of acute rejection, for whom the benefits of the stronger treatment B are limited, now experience net harm from treatment B, while those at high risk for acute rejection, with a greater opportunity for benefit, still experience net benefit from treatment B. Summary results of such a trial would report overall null effects, and conventional subgroup analyses are likely to be uninformative. Risk stratified analyses,

however, are much more likely to identify the high risk subgroup to which the stronger, but more nephrotoxic regimen should be targeted.

## Identification of Risk Groups of Patients

Exploring treatment effect heterogeneity in kidney transplant recipients, stratified according to their baseline risk for rejection would be a helpful guide to identify patients who benefit most from a particular immunosuppressive regimen. Although the underlying pathophysiological processes and genetic underpinnings are incompletely understood, various factors have been identified that describe “immunological risk”, such as African American race, panel reactive antibody (PRA) level, previous transplantation, and human leukocyte antigen (HLA) mismatch. (21–24) A predictive model developed with these variables is likely to achieve sufficient discrimination to form the basis of an informative risk-based analyses.

While differences in the outcome of interest are likely to lead to differences in the risk-benefit trade-offs of different immunosuppressive regimens, two other dimensions of risk can also underlie clinically important treatment effect heterogeneity and warrant further study: differences in the risk of treatment-related harms and differences in competing risks. Stronger immunosuppressive regimens clearly reduce the risk for rejection but also increase the risk for infections and malignancies.(25) Little is known about other possible predisposing factors and whether there is important heterogeneity in baseline risk. It is also well appreciated that many patients die with a functioning graft, primarily due to cardiovascular disease. (15;25;26) Differences in the risk of cardiovascular disease, which may limit a patient’s likelihood of benefiting from successful transplantation may also underlie differences in the probability of benefiting from different immunosuppressive regimens. This is particularly true when these agents worsen hypertension, dyslipidemia and diabetes.

## Sources for the Development of Prediction Models

Large observational databases and registries are potential sources to develop prediction models for acute rejection, since many of the predictors of immunological risk are known and commonly collected. However, these datasets frequently lack information about CVD events, infections and some potentially important risk factors. Prospective cohort studies with sufficiently large sample sizes and long durations of follow-up are currently not available, but are on the way.(27) Another source of data could be developed by combining several RCT datasets and performing the analyses in a patient-level meta-analysis. This approach has the advantage that patients are characterized in detail, along with multiple clinical and important safety outcomes. However, most RCTs exclude “high-risk” patients and long-term events, such as malignancies, CVD events and death, are commonly not captured.

Conventional meta-analysis would not be helpful for this purpose. Meta-analyses integrate evidence from average findings of published RCTs in an overall effect estimate. Based on the underlying enrollment criteria, the study populations can vary dramatically, which is particularly common in trials of induction agents (28) Standard meta-analysis would further dilute any differences in outcomes due to patients’ underlying risks. While subgroup meta-analyses can be performed, these would suffer from the same limitations as within-study subgroup analyses. Meta-regression techniques have been shown to be insufficient for detecting differences in treatment effect due to patient-level variables.(29)

## The Impact of Risk-Based Analyses on Future Research

When a risk stratification model is available, patients can be stratified according to their immunological risk. For example, a 35 year old sensitized patient with a PRA level of 70%, receiving an allograft from a living donor with a positive cross-match could be classified as high risk for acute rejection. A 65 year old hypertensive diabetic patient with a history of CVD receiving a graft with no HLA mismatches and a cold ischemia time of 12 hours could be considered as low risk for rejection, however, with a high risk for further CVD events.

Following this approach, existing RCTs could be re-analyzed and treatment effect heterogeneity could be explored. However, since many “high risk” patients are commonly excluded from RCTs, evidence from these post hoc analyses would substantially underestimate the range of variability of risk seen in clinical practice. Nevertheless, the analyses could be helpful to explore the risk-modeling approach of identifying subgroups with different risks and benefits of more intensive immunosuppression. If these tools show promising results and are validated in other observational cohorts, registries, and existing trial databases, the proposed risk categories should be investigated prospectively in RCTs. A priori defined subgroups of patients according to their baseline risk could be implemented in the randomization process and patients allocated to varying treatments, such as stronger immunosuppressive regimens for “high risk” patients and standard regimens for “low risk” patients. This promising design has recently been implemented in a trial in which alemtuzumab was tested against thymoglobulin in “high risk” patients or basiliximab in “low risk” patients.(30)

The goal of this approach is to allow clinicians to tailor treatment to a patient’s baseline risks: strong immunosuppression for patients at high risk of rejection, but less strong immunosuppression for patients at lower risk of rejection who may be more vulnerable to treatment-related harms. The ability to determine who is at increased risk for different outcomes and tailoring treatment has the potential to result in overall improved patient and graft survival and reduce adverse events. Although simulations in the setting of kidney transplantation have not been performed, it is conceivable that considering risk strata would not dramatically inflate the sample size in RCTs to achieve significant findings.(10)

## Conclusion

Average results of RCTs often fail to describe particular treatment effects based on varying baseline risk. Although challenging to establish, risk-based analyses are a promising approach to explore treatment effect heterogeneity in kidney transplantation. Once the risk models are validated they may be useful in the planning of clinical trials and in their analysis. Ultimately, this approach should help match the best immunosuppressive regimens for individual patients.

## Acknowledgments

**Funding Source/Acknowledgements:** Dr. Wagner receives funding from the fellowship training program of the National Kidney Foundation Center for Clinical Practice Guideline Development and Implementation at Tufts Medical Center. Dr. Balk receives funding from the National Kidney Foundation Center for Clinical Practice Guideline Development and Implementation at Tufts Medical Center. Dr. Kent is supported by a grant of the National Institute of Health (NIH/NCRR 1UL1 RR025752)

## Reference List

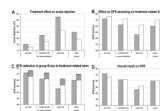
1. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001 Apr 14;357(9263): 1191–4. [PubMed: 11323066]



2. Schold JD, Kaplan B. Design and Analysis of Clinical Trials in Transplantation: Principles and Pitfalls. *Am J Transplant*. 2008 Jul 28;
3. Rothwell PM. Can overall results of clinical trials be applied to all patients? *Lancet* 1995 Jun 24;345(8965):1616–9. [PubMed: 7783541]
4. Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004;82(4):661–87. [PubMed: 15595946]
5. Ioannidis JP, Lau J. Heterogeneity of the baseline risk within patient populations of clinical trials: a proposed evaluation algorithm. *Am J Epidemiol* 1998 Dec 1;148(11):1117–26. [PubMed: 9850135]
6. Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007 Sep;Dec;298(10):1209–12. [PubMed: 17848656]
7. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000 Mar 25;355(9209):1064–9. [PubMed: 10744093]
8. Oxman AD, Guyatt GH. A consumer's guide to subgroup analyses. *Ann Intern Med* 1992 Jan 1;116(1):78–84. [PubMed: 1530753]
9. Jassal SV, Schaubel DE, Fenton SS. Baseline comorbidity in kidney transplant recipients: a comparison of comorbidity indices. *Am J Kidney Dis* 2005 Jul;46(1):136–42. [PubMed: 15983967]
10. Hayward RA, Kent DM, Vijan S, Hofer TP. Multivariable risk prediction can greatly enhance the statistical power of clinical trial subgroup analysis. *BMC Med Res Methodol* 2006;6:18. [PubMed: 16613605]
11. Kent DM, Hayward RA, Griffith JL, Vijan S, Beshansky JR, Califf RM, et al. An independently derived and validated predictive model for selecting patients with myocardial infarction who are likely to benefit from tissue plasminogen activator compared with streptokinase. *Am J Med* 2002 Aug 1;113(2):104–11. [PubMed: 12133748]
12. Kent DM, Jafar TH, Hayward RA, Tighiouart H, Landa M, de JP, et al. Progression risk, urinary protein excretion, and treatment effects of angiotensin-converting enzyme inhibitors in nondiabetic kidney disease. *J Am Soc Nephrol* 2007 Jun;18(6):1959–65. [PubMed: 17475813]
13. Vijan S, Kent DM, Hayward RA. Are randomized controlled trials sufficient evidence to guide clinical practice in type II (non-insulin-dependent) diabetes mellitus? *Diabetologia* 2000 Jan;43(1):125–30. [PubMed: 10672454]
14. U.S. Food and Drug Administration. Xigris Product Information. 2009. [http://www.fda.gov/medwatch/safety/2009/Xigris\\_PI.pdf](http://www.fda.gov/medwatch/safety/2009/Xigris_PI.pdf)
15. Tantravahi J, Womer KL, Kaplan B. Why hasn't eliminating acute rejection improved graft survival? *Annu Rev Med* 2007;58:369–85. [PubMed: 17002551]
16. Meier-Kriesche HU, Ojo AO, Hanson JA, Cibrik DM, Pirsch JD, Leichtman AB, et al. Increased impact of acute rejection on chronic allograft failure in recent era. *Transplantation* 2000 Oct 15;70(7):1098–100. [PubMed: 11045649]
17. A blinded, randomized clinical trial of mycophenolate mofetil for the prevention of acute rejection in cadaveric renal transplantation. The Tricontinental Mycophenolate Mofetil Renal Transplantation Study Group. *Transplantation* 1996 Apr 15;61(7):1029–37. [see comment]. [PubMed: 8623181]
18. Mathew TH. A blinded, long-term, randomized multicenter study of mycophenolate mofetil in cadaveric renal transplantation: results at three years. Tricontinental Mycophenolate Mofetil Renal Transplantation Study Group. *Transplantation* 1998 Jun 15;65(11):1450–4. [PubMed: 9645801]
19. Hazzan M, Buob D, Labalette M, Provot F, Glowacki F, Hoffmann M, et al. Assessment of the risk of chronic allograft dysfunction after renal transplantation in a randomized cyclosporine withdrawal trial. *Transplantation* 2006 Sep 15;82(5):657–62. [PubMed: 16969289]
20. Ekberg H. Graft survival benefit to be expected of new immunosuppressive regimens. *Transplantation Reviews* 2003;17(4):187–93.
21. Aydingoz SE, Takemoto SK, Pinsky BW, Salvalaggio PR, Lentine KL, Willoughby L, et al. The impact of human leukocyte antigen matching on transplant complications and immunosuppression dosage. *Hum Immunol* 2007 Jun;68(6):491–9. [PubMed: 17509448]

22. Kerman RH, Kimball PM, Van Buren CT, Lewis RM, Kahan BD. Possible contribution of pretransplant immune responder status to renal allograft survival differences of black versus white recipients. *Transplantation* 1991 Feb;51(2):338–42. [PubMed: 1825242]
23. Sanfilippo F, Vaughn WK, LeFor WM, Spees EK. Multivariate analysis of risk factors in cadaver donor kidney transplantation. *Transplantation* 1986 Jul;42(1):28–34. [PubMed: 3523879]
24. Thibaudin D, Alamartine E, de Filippis JP, Diab N, Laurent B, Berthoux F. Advantage of antithymocyte globulin induction in sensitized kidney recipients: a randomized prospective study comparing induction with and without antithymocyte globulin. *Nephrol Dial Transplant* 1998 Mar; 13(3):711–5. [PubMed: 9550651]
25. Pascual M, Theruvath T, Kawai T, Tolkoff-Rubin N, Cosimi AB. Strategies to improve long-term outcomes after renal transplantation. *New England Journal of Medicine* 2002 Feb 21;346(8):580–90. [see comment]. [Review] [99 refs]. [PubMed: 11856798]
26. Halloran PF. Immunosuppressive drugs for kidney transplantation.[erratum appears. *New England Journal of Medicine* 2004 Dec 23;351(26):2715–29. [erratum appears in *N Engl J Med*. 2005 Mar 10;352(10):1056]. [Review] [124 refs]. [PubMed: 15616206]
27. Kasiske BL, Israni AK, Snyder JJ, Skeans MA, Weinhandl ED, Peng Y. Cardiac Events Post Kidney Transplantation: Interim Results from the Patient Outcomes in Renal Transplantation (PORT) International Transplant Study. *Am J Transplant* 2008;8(s2):177–336. Ref Type: Abstract. [PubMed: 18435674]
28. Webster AC, Playford EG, Higgins G, Chapman JR, Craig J. Interleukin 2 receptor antagonists for kidney transplant recipients. *Cochrane Database Syst Rev* 2004;(1):CD003897. [PubMed: 14974043]
29. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004 Jul;57(7):683–97. [PubMed: 15358396]
30. Hanaway M, Woodle SE, Mulgaonkar S, Peddi R, Harrison G, Vandeputte K, et al. 12 Month Results of a Multicenter, Randomized Trial Comparing Three Induction Agents (Alemtuzumab, Thymoglobulin and Basiliximab) with Tacrolimus, Mycophenolate Mofetil and a Rapid Steroid Withdrawal in Renal Transplantation. [abstract]. *Am J Transplant* 2008;8(s2):177–336. Ref Type: Abstract. [PubMed: 18435674]





### Figure 1. How risk stratification can explain the paradox in the hypothetical RCT

Illustration of the paradox discussed in the text about the hypothetical randomized trial described in the table. Treatment A in grey columns and treatment B in white columns, respectively. Panel A: Acute rejection rates associated with both treatments, with a constant relative risk reduction of 40% in the entire cohort as well as in subgroups stratified by baseline risk for acute rejection. Panel B: If acute rejection represents the only predictor of worse graft outcome, treatment B is associated with an improved glomerular filtration rate (GFR) due to the reduced rate of acute rejection. The actual effect size is dependent on the patient's baseline risk for acute rejection. Panel C: We assume that only treatment B is associated with nephrotoxic effects that reduce GFR by 10 ml/min (striped boxes), and that this is constant across risk strata. Panel D: Overall treatment effect on GFR, considering treatment-related adverse events. (Detailed explanation in the text)

**Table 1**

Hypothetical randomized controlled trial results

	Group A N=200	Group B N=200
Acute rejection, N (%)	60 (30 %)	36 (18 %)
Mean glomerular filtration rate	60 ml/min	60 ml/min
Number of patients with decreased glomerular filtration rate due to acute rejection*	60	36

\* Acute rejection as a strong risk factor for impaired kidney function