Appendix 1: Assembly of study dataset and detailed methods for analysis

To protect the privacy of subjects, the Canadian Organ Replacement Registry (CORR) initially provided us with a 2-variable dataset containing the 6-digit residential postal code as recorded at the time of dialysis initiation, together with an encrypted identification number. These data were provided for a random 75% sample of all patients initiating dialysis in Canada between Jan. 1, 1996, and Dec. 31, 2000 (n = 15 044). Of these patients, the residential postal code was valid for 14 541 (96.7%). We calculated the geographic coordinates for each patient's residence using the Canadian postal code conversion file,[1] and determined the transplant centre that would provide care to each study subject on the basis of each centre's defined catchment area. We then calculated the distance from the residence location to the transplant centre for all 14 541 participants (described below) and returned these data to CORR together with the encrypted identification number. CORR then linked these geographic data to demographic and clinical data, stripped all uniquely identifying information from the file and randomly selected 7151 (about 50%) of the original subjects. After excluding children (age < 18 years, n = 117), we performed analyses on the resulting dataset, which included clinical and demographic data, geographical location and distance from transplant centre for 7034 patients (about 36% of all subjects initiating dialysis in Canada during the study period).

Estimation of distance and travel time

The geographic coordinates for each 6-digit postal code were determined using the Statistics Canada Postal Code Conversion File and Canadian*PostalData software (Melissa Data Corporation, Rancho Santa Margarita, Calif.). These coordinates were entered into ArcGIS 3.0 software (ESRI Corporation, Redlands, Calif.) to determine the shortest distance by road (in kilometres) between the residence of a dialysis patient at the time of dialysis initiation and the facility that would perform his or her renal transplant or transplantation workup as previously described,[2-4] or both.

Estimation of physician supply

Data from the Southam Medical Database were used to determine the population:physician ratio in each census consolidated subdivision (CCS) during the year in which each patient initiated dialysis.[5] We estimated socioeconomic status using the neighbourhood income per person equivalent (IPPE), a household size-adjusted measure of household income, based on 1996 Canadian census summary data.[1]

Statistical analysis

The adjusted association between residence location and time to transplantation was determined using a Cox proportional hazards model. Other factors considered included age, sex, race (white, Aboriginal, other or unknown); cause of end-stage renal disease; year of dialysis initiation; comorbid conditions, including diabetes mellitus, coronary disease, previous or current hypertension, chronic heart failure, stroke or transient ischemic attack, chronic lung disease, peripheral vascular disease, malignant disease; smoking status; initial mode of dialysis; socioeconomic status; and population:physician ratios. We also adjusted for the annual rate of kidney donation from deceased donors in each region (per million population), which was entered into the models as a time-varying covariate (updated annually to reflect changes in donation rate) with a time-dependent coefficient (to allow the effect of the donor rate to depend on time since initiation of dialysis). To determine whether the effects of residence location differed within each geographic region, we explored models stratified by geographic region. We determined that the proportional hazard assumption was satisfied by examining plots of the log-negative-log of the within-group survivorship probabilities versus log-time and by comparing Kaplan–Meier (observed) and Cox (estimated) survival curves.

We dealt with missing data by assuming that the characteristic was absent (11% had missing data on one or more comorbid conditions) or by representing the missing data with an indicator variable (13% had missing data on race). Results did not differ when analyses were repeated after deleting all subjects with missing data, so we have reported results using the former method. To deal with other methodological uncertainties, we conducted the following sensitivity analyses. We also categorized distance into 6 categories (< 50 km; 50.1–100 km; 100.1–200 km; 200.1–400 km; 400.1–600 km; and > 600 km) as well as the 4 categories in the primary analysis (< 50 km; 50.1–150 km; 150.1–300 km; and > 300 km). Because we did not have information on transplant wait-list status, it is possible that comorbid disease affected eligibility and thus rates of transplantation. We addressed this with an additional analysis in which we considered only subjects aged less than 60 years without known diabetes mellitus, coronary disease, chronic heart failure, stroke, chronic lung disease, peripheral vascular disease, known malignant disease or other serious medical illness that might reduce the likelihood of wait-listing for transplantation. We also examined the effect of including transplants from living donors. Finally, we assessed the potential for informative censoring (since those who die early are less likely to receive a kidney transplant) by assuming that patients who died would not have received a transplant if they had survived until the end of the study period (i.e., patients who died were assigned a date of last follow-up of Dec. 31, 2002).

We used logistic regression analysis to estimate the predicted proportion of patients receiving a kidney transplant from a deceased donor in the first 3 years following initiation of dialysis. We also examined differences in expected waiting time between geographic regions in hypothetical groups of patients characterized by age and diabetic status. We estimated predicted median time to transplantation in these groups (i.e., the time until 50% of all patients received a transplant) using a parametric model[6] that assumed a log-logistic distribution for the hazard function.

References

1. Ng E, Wilkins R, Perras A. How far is it to the nearest hospital? Calculating distances using the Statistics Canada Postal Code Conversion File. *Health Rep* 1993;5:179-88.
2. Larson RC. *Urban operations research*. Englewood Cliffs (NJ): Prentice-Hall; 1981.
3. Brabyn L, Gower P. Comparing three GIS techniques for modeling geographical access to general practitioners. *Cartographica* 2004;39:41-9.
4. Brabyn L, Skelly C. Modeling population access to New Zealand public hospitals. *Int J Health Geogr* 2002;1:3.
5. Pitblado JR, Pong RW, Irvine A, et al. *Assessing rural health: toward developing health indicators for rural Canada*. Ottawa: Health Canada; 1999. Available: www.laurentian.ca/cranhr/pdf/indcat/INDCONT.pdf (accessed 2006 Jun 6).
6. Lawless JF. Parametric regression models. In: *Statistical models and methods for lifetime data*. 2nd ed. Hoboken (NJ): John Wiley and Sons; 2002. p. 269-330.

Appendix to Tonelli M, Klarenbach S, Manns B, et al; Alberta Kidney Disease Network. Residence location and likelihood of kidney transplantation. *CMAJ* 2006;175(5):478-82.