Investigation of measurement invariance in longitudinal health-related quality of life in preemptive or previously dialyzed kidney transplant recipients

Line Auneau-Enjalbert, Myriam Blanchin, Magali Giral, Aurélie Meurette, Emmanuel Morelon, Laetitia Albano, Jean-Benoit Hardouin, Véronique Sébille

Investigation of measurement invariance in longitudinal health-related quality of life in preemptive or previously dialyzed kidney transplant recipients

Line Auneau-Enjalbert[1,2], Myriam Blanchin[1], Magali Giral[3*], Aurélie Meurette[3*], Emmanuel Morelon[4*], Laetitia Albano [5]*, Jean-Benoit Hardouin[1,2], Véronique Sébille[1,2]

[1]: UMR INSERM 1246, SPHERE "methodS in Patient-centered outcomes and HEalth ResEarch", Nantes University, Tours University, Nantes, France;

[2]: Methodology and Biostatistics unit, CHU Nantes, Nantes, France ;

[3]: CRTI UMR 1064, Inserm, Université de Nantes; ITUN, CHU Nantes; RTRS Centaure, Nantes, France;

[4]: Néphrologie, Transplantation et Immunologie Clinique, Hôpital Edouard Herriot, Lyon, France;

[5]: Service de Néphrologie, Hôpital Pasteur, Nice, France

** Données Informatisées et VAlidées en Transplantation, DIVAT Cohort Collaborators listed at end of manuscript*

Corresponding author:

Véronique Sébille, UMR INSERM 1246, SPHERE "methodS in Patient-centered outcomes and HEalth ResEarch" Nantes University, Tours University, 22 boulevard Benoni Goullin, Nantes, France. Phone number: 0240412825. Email: veronique.sebille@univ-nantes.fr

Acknowledgments

The study has received ethics approval and is supported by a grant from the French Ministry of Health (PHRC-13-0224, 2013). The authors express sincere thanks to Elodie Faurel-Paul and Astrid Fleury for their assistance for the study, as well as all the participants for their contribution to the study. The authors wish to thank members of the clinical research assistant team and DIVAT Consortium Collaborators (Medical Doctors, Surgeons, HLA Biologists)

Nantes : Gilles Blancho, Julien Branchereau, Diego Cantarovich, Agnès Chapelet, Jacques Dantal, Clément Deltombe, Lucile Figueres, Claire Garandeau, Magali Giral, Caroline Gourraud-Vercel, Maryvonne Hourmant, Georges Karam, Clarisse Kerleau, Aurélie Meurette, Simon Ville, Christine Kandell, Anne Moreau, Karine Renaudin, Anne Cesbron, Florent Delbos, Alexandre Walencik, Anne Devis; Lyon E. Hériot : Lionel Badet, Maria Brunet, Fanny Buron, Rémi Cahen, Sameh Daoud, Coralie Fournie, Arnaud Grégoire, Alice Koenig, Charlène Lévi, Emmanuel Morelon, Claire Pouteil-Noble, Thomas Rimmelé, Olivier Thaunat.

Abstract

Purpose

Kidney transplantation (KT) can impact patients' evaluation of health-related quality of life (HRQoL) as they adapt to their new life with a graft and its changes. Patients may adapt to KT in a different way, depending on whether they were on dialysis prior to transplantation or not (i.e. preemptive group). This may result in lack of measurement invariance between these patients' groups and/or over time (i.e. response shift, RS) which may invalidate the between-group comparison of HRQoL change scores. The aim of this study was to investigate and compare RS before and after KT between these two patients' groups. Measurement invariance was investigated between groups and over time with three measurement occasions.

Methods

Adult patients completed the SF-36 at the last visit before KT, and 3, 6 months after. A structural equation model-based procedure was used to (i) detect and take into account measurement non-invariance between groups and RS, if appropriate, (ii) identify the period of occurrence of RS, (iii) study the heterogeneity of RS between the two groups.

Results

Before KT (i.e. baseline), measurement invariance was not rejected between dialyzed (n = 196) and preemptive (n = 178) patients' groups. Between baseline and 3-month after KT, similar uniform recalibration was detected on the general health domain in both groups. Uniform recalibration was found between 3- and 6-month after KT on the vitality domain for preemptive patients only.

Conclusion

HRQoL, adjusted for RS, increased overall for preemptive and dialyzed kidney transplant patients after transplantation. RS may reflect differing adaptation processes following KT.

Keywords: structural equation model, response shift, measurement invariance, longitudinal data, kidney transplantation

Introduction

Treatment of end-stage renal disease by dialysis or kidney transplantation (KT) can have an impact on the physical and psychological health of patients, including their health-related quality of life (HRQoL). When it is possible, KT remains the treatment of choice because it has many benefits compared to dialysis. Indeed, several studies have shown that kidney transplant patients have a better long-term survival and lower risks of cardiovascular events [1, 2], as well as a better HRQoL [3–5] compared to patients undergoing dialysis. Furthermore, preemptive transplantation (i.e. without a period of dialysis before transplantation) would be the preferred strategy to avoid trauma associated with dialysis. Whether it be a preemptive transplant or not, patients have to adapt to their new life with a graft and the changes that result from it. We can expect that the HRQoL of preemptive and dialyzed transplant patients may be different over time, before and after KT, depending on their previous experience with renal replacement therapy (i.e. dialysis or not) and expectations. Only a few studies have compared HRQoL changes of preemptive and dialyzed kidney transplant patients before and after transplantation [6–8] and most have been realized on small sample sizes. In all these studies, the patient's perception of HRQoL was assumed to be the same between patient groups and over time, i.e. assumption of measurement invariance [9]. If this assumption does not hold, observed HRQoL score changes or difference of observed scores between groups may not properly reflect change or differences in the latent construct (e.g. HRQoL).

When the assumption of measurement invariance is invalid with respect to time, it is a special case of longitudinal measurement non-invariance, also named response shift (RS). RS refers to a change in the meaning of one's self-evaluation of a target construct that results from a change in one's internal standards (recalibration), a change in one's values (reprioritization), or one's redefinition of the target construct (reconceptualization) [10]. RS is assumed to occur in particular after a serious health event, also named catalyst (e.g. KT in our case), and may

often be related to how patients adapt to their experience of the disease [11]. It is important to detect and account for RS to ensure a valid assessment of change in patients' HRQoL (i.e. target change) over time and to avoid possibly erroneous conclusions. Moreover, RS is important to study per se as it may also provide a better understanding of how patients adapt to their illness.

Several statistical methods have been developed to assess and take into account lack of measurement invariance between groups, over time, or both [12, 13], including structural equation modelling (SEM) [14, 15]. SEM is a versatile method which allows exploring the different types of RS (i.e. recalibration, reprioritization, and reconceptualization) and adjusting for measurement non-invariance between groups and over time, if appropriate [16]. Changes in the following SEM parameters are assumed to be indicative of RS: intercepts (uniform recalibration), factor loadings (reprioritization), and residual variances (non-uniform recalibration). In longitudinal studies, HRQoL questionnaires are often collected at more than two occasions, and researchers have sought to detect possible RS effects over multiple occasions, also considering the heterogeneity of the studied population with multiple-group approaches. Verdam and Oort [17] have applied the longitudinal three-mode model to assess HRQoL change of patients with painful bone metastasis before and after radiotherapy treatment, over many measurement occasions. These models were also extended to multiple-group analyses [18]. However, in these models, an overall trend of change in SEM parameters is obtained, without providing information with regards to the period where RS may have occurred, or not.

Our aim is to investigate measurement invariance in longitudinal HRQoL before and soon after KT in preemptive and dialyzed transplant patients. Two consecutive periods were considered: before and just after KT (at 3 months) and between 3 and 6 months. These two time periods, very close to surgery, are considered critical for patients from a clinical and psychological point of view [19, 20]. Our objective was to investigate RS and to get more

insight into patients' adaptation to KT in each of these two time periods using an adapted SEM-based approach. This paper proposes a SEM-based procedure to: (1) detect and take into account measurement non-invariance between and within groups across more than two measurement occasions, (2) identify the period of occurrence of RS, and (3) study the heterogeneity of RS between the two groups.

Methods

Study population

In the PreKit-QoL study [21], patients are recruited from three centers participating in the DIVAT cohort (www.divat.fr, approved by the Comité National de l'Informatique et des Libertés CNIL, no 914184) in France: Nantes, Nice, and Lyon. Adult patients (over 18 years old) registered on the waiting list for first KT without other transplanted organs, who have given their informed consent are prospectively pre-included since September 2014. The PreKit-QoL study is registered on the ClinicalTrials.gov Registry (RC14_0078, NCT02154815), has obtained approval from the ethical Committee for Persons' Protection (CPP, Tours, 2014-S8), and from the advisory committee on research data and information in health (CCTIRS, Paris, 14.314).

Pre-included patients are followed on the waiting list until they receive KT or experience another event (dialysis>36 months, death, study withdrawal). Transplanted patients are included in the PreKit-QoL study and two groups are distinguished: patients receiving KT after a period of dialysis<36 months, and patients receiving KT without a dialysis experience (preemptive patients).

Measures

HRQoL and psychological data are collected in the two groups, every 6 months before transplantation, at the hospital discharge, and after transplantation, at 3 months, 6 months, and every year until 5 years of follow-up. For the present study, we focus on the last visit before KT (baseline) and the 3- and 6-month visits after KT.

HRQoL is measured using the SF-36 [22], a generic and validated questionnaire composed of 8 domains (Physical Functioning, PF; Role Physical, RP; Bodily Pain, BP; General Health, GH; Vitality, VT; Social Functioning, SF; Role Emotional, RE; and Mental Health, MH). The scores of each domain range from 0 to 100, a higher score indicating a better HRQoL.

Conceptual overview of the rationale behind the methodological framework

The first part of the procedure is designed to test whether full or partial measurement invariance holds between dialyzed and preemptive groups at baseline. If full invariance is rejected, this may suggest that patients have differing perceptions of their HRQoL at baseline, depending on whether they experienced dialysis or not. The second part is based on SEM following Oort [14] and on the definition of RS as a violation of the principle of conditional independence in measuring latent variables over time [23, 24]. In this framework, RS is assumed to have occurred when observed change is not fully explained by target change due to, e.g., change in factor loadings (reprioritization).

Methodological framework

The procedure is composed of a series of SEM fitted using Stata software version 15.0 (StataCorp, College station, TX). Missing data for the SF-36 domains scores are handled using the estimation method of full information maximum likelihood. Homogeneity of RS is assumed within each patient group.

The evaluation of the goodness of fit for each model is assessed using the ratio χ^2/df, where df is the degree of freedom, the root mean square error of approximation (RMSEA) and comparative fit index (CFI) [25, 26]. A value below 0.08 for the RMSEA [27] and above 0.90 for the CFI indicates an acceptable fit [28]. A likelihood ratio test (LRT) based on the difference of chi-square statistics between constrained and unconstrained models is used to compare nested models. The procedure described in figure 1, is composed of 2 parts, each including several steps. It is described for two groups (preemptive and dialyzed patients) and three measurement occasions (baseline, 3, and 6 months after KT).

Before starting the procedure, a preliminary analysis is made using the clustering around latent variables method [29] to determine the structure of the SF-36 questionnaire (i.e., which domain scores are indicative of which latent variables), based on our data. Clustering around latent variables consists in a hierarchical cluster analysis that assigns each observed variable to the latent variable with which it is the most correlated. This provides the measurement model used in part 1 described below, assumed to be the same for preemptive and dialyzed patients.

Part 1: Detection of measurement non-invariance between groups at baseline

Step A: Measurement model with non-invariance between groups

In this cross-sectional multigroup SEM model, all factor loadings, intercepts, and residual variances are freely estimated between the two groups (i.e. unconstrained model). Model modifications for the improvement of model fit are guided by modification indices (also named Lagrangian multiplier test [30]) and assessed using a LRT.

Step B: Measurement model with invariance between groups and overall test of invariance

The unconstrained model of step A is compared to a fully constrained model using a LRT. If the test is not significant at the 5% significance level, invariance between groups is assumed at baseline and we continue with part 2 of the procedure for the detection of RS (step 1). Otherwise, we proceed with step C.

Step C: Identification of domains with non-invariance between groups

From step B, in the first iteration, score and Wald tests are used to test parameters' invariance between groups on each domain. The equality constraint of the parameter with the highest chi-square statistic is relaxed between groups for the identified domain, and the model is compared to the constrained model using a LRT. If the test is significant at the 5% significance level, invariance of the parameter is rejected, and the model is updated, with the identified constraint relaxed. The same steps are continued iteratively until the LRT is no longer significant i.e. partial measurement invariance between groups hold, or when measurement non-invariance between groups have been detected on D-1 domains among the D domains.

Part 2: RS detection

Step 1: Measurement model with RS

The last model obtained in part 1 is extended to include a second measurement occasion. In this longitudinal multigroup SEM, the factor loadings, intercepts, and residual variances of all domains are freely estimated between baseline and 3 months after KT, and between groups at the second measurement occasion (i.e. unconstrained model with RS).

Step 2: No RS model and overall test of RS

A no RS model, where the model parameters are constrained to be equal over time (i.e. fully constrained model), is estimated and compared to the unconstrained model of step 1 using a

LRT. If the LRT is not significant at 5% significance level, no RS is assumed between the baseline and 3-month visit after KT, and we continue RS detection between the 3- and 6-month visits. Otherwise, we proceed with step 3.

Step 3: Identification of RS

This step aims to identify the type of RS, and which domains and groups are impacted. From the step 2 model, the equality constraint of the domain with the highest modification indices in the specified group is relaxed. This model is compared to the constrained model using a LRT. If the test is significant at 5% significance level, RS in the specified group on the identified parameter is assumed. The iterations continue until relaxing constraints does not improve the model further (i.e. the LRT is no longer significant) or when RS has been detected on D-1 domains among D domains. When an equality constraint is relaxed in both groups, we test whether RS occurs in the same way (similar RS) or not (differential RS) between groups using a Wald test.

RS detection for more than two occasions

The procedure continues by adding a third measurement occasion (6-month visit after the KT). Steps 1 to 3 are repeated to assess RS between the 3- and 6-month visits after KT, in order to test if the assumption of measurement invariance is tenable taking into account the results found previously. The final model of step 3 is used to estimate the magnitude of RS effects between measurement occasions in each patient group, using effect-size indices, where values of 0.2, 0.5 and 0.8 are considered small, medium and large, respectively [14, 31].

Step 4: Target change assessment

The change in patient's HRQoL (i.e. target change) is assessed from the final model of step 3 taking into account the non-invariance of parameters detected in the previous steps, if appropriate. The group and time effects on the latent construct and their interaction are assessed using Wald tests.

Identifiability constraints

The model's identifiability is achieved by constraining the means of the latent variables to zero and variances to 1 in each group at the baseline measurement occasion in step A, and in the dialyzed patients group only for steps B, C, and 1 to 3. From step 1, the means and variances in preemptive and dialyzed patients' groups are constrained to be equal over time.

Results

In the PreKit-QoL study, 374 patients have received a kidney transplant since September 2014, including 196 patients in the dialyzed group (52%) and 178 patients in the preemptive group (48%). Sociodemographic and transplantation associated clinical and biological variables are described for the two groups at baseline (Table 1). These variables are similar in both groups, except for more diabetes comorbidity in the dialyzed patients group compared to the preemptive group. Missing data, observed scores means and their standard deviations, and estimates of skewness and kurtosis are reported in Supplementary Table A for each SF-36 domains in each group and at each measurement occasion.

The measurement model is determined using the clustering around latent variables, based on the 8 domains of the SF-36 questionnaire [22] and our data. The model comprises 2 latent variables composed of 4 domains each. The latent variable measured by the PF, RP, BP, and

RE domains is named physical HRQoL and the latent variable measured by SF, VT, MH and GH is named mental HRQoL.

Part 1: Detection of measurement non-invariance between preemptive and dialyzed patients at baseline

Step A: Measurement model with non-invariance between preemptive and dialyzed patients

For computational convenience, scores of the SF-36 questionnaire are divided by 100. For physical HRQoL, correlations between residual variances of RP, RE, and BP, RE are added to improve model fit. Correlations between residual variances of SF, MH, and VT, GH are included for mental HRQoL. The final measurement model had a reasonable fit (χ^2/df = 2.15, RMSEA = 0.078, CFI = 0.972, Table 2, Figure 2).

Step B: Measurement model with invariance between preemptive and dialyzed patients and overall test of differences between groups

The LRT comparing a constrained model with the previous unconstrained model (Table 2) is not significant (p = 0.096). Factor loadings, intercepts, and residual variances are therefore constrained to be equal between preemptive and dialyzed groups at baseline.

Part 2: RS detection between the baseline and 3-month visit after KT

Step 1: Measurement model with RS

The previous multigroup model is extended to a longitudinal model with RS between baseline and 3-month visit after KT, in the preemptive and dialyzed patient groups.

Step 2: No RS model and overall test of RS

The LRT comparing the no RS model to the RS model of step 1 (Table 2) is significant ($p < 0.001$), indicating the occurrence of RS between baseline and 3-month visit after transplantation.

Step 3: Identification of RS

Constraints on intercepts of the GH domain are relaxed in the preemptive and in the dialyzed patients' groups, suggesting the occurrence of uniform recalibration RS. The Wald test indicates that RS occurs in the same way in both groups ($p = 0.99$) indicating a similar uniform recalibration. Intercept estimates (Table 3) suggest that, preemptive and dialyzed patients tended to report a better GH (approximately +6 points) at 3-month visit after KT (approximately 52 points) compared to baseline (approximately 46 points) given similar mental HRQoL. No other RS effect has been identified on the other domains.

RS detection between 3- and 6-month visits after KT

From the final model of step 3, a model with RS is estimated (step 1) and compared to a no RS model (step 2) using a LRT. The LRT (Table 2) is significant ($p = 0.039$), indicating the occurrence of RS across these 2 visits. The parameter estimates for the final model are presented in Table 3 and the correlations between latent variables are presented in supplementary Table B.

In step 3, the constraint on intercepts for the VT domain is relaxed in the preemptive patients' group only suggesting the occurrence of uniform recalibration. The intercept estimates (Table 3) of the VT domain suggest that, preemptive patients tended to report a better vitality (approximately +3 points) at 6-month visit after KT (approximately 48 points) compared to 3-month visit (approximately 45 points) given similar mental HRQoL. No other RS effect has been identified on the other domains between these 2 visits.

RS effect sizes are estimated on the final model of step 3, between baseline and 3-month after KT, and 3- and 6- months after KT, in the preemptive and dialyzed groups(Table 4). A small uniform recalibration effect size is observed for the GH domain in preemptive and dialyzed patients' groups, between baseline and 3-months after KT. A small uniform recalibration effect size is also observed for the VT domain in the preemptive group only, between 3- and 6-months after KT.

Step 4: Target change assessment

Preemptive patients have on average a significant better physical HRQoL than patients in the dialysis group at baseline (+ 0.33, $p = 0.006$), with similar mental HRQoL ($p = 0.98$) (Table 5). The physical HRQoL (Figure 3) in the preemptive group remains stable before and 3 months after KT ($p = 0.78$), and a significant improvement between 3 and 6 months is evidenced after transplantation (+ 0.37, $p < 0.001$). For the dialyzed patients' group, the physical HRQoL significantly increases over time (+ 0.20 between baseline and 3-month visits after transplantation, and + 0.25 between 3- and 6-month visits, $p = 0.004$). No RS was evidenced hence, estimation of the means of the physical HRQoL are not affected by RS (Figure 3).

The mental HRQoL (Figure 4) improves significantly in preemptive and dialyzed patients' groups between baseline and 3-month visit after transplantation (+ 0.63 and + 0.61, respectively, $p < 0.001$ for both groups), and then remains stable until 6 months' post-transplantation ($p = 0.96$ and $p = 0.38$, respectively). Although, RS was evidenced for mental HRQoL, estimated means were only slightly affected by RS (Figure 4).

Discussion

In this study, measurement invariance was not rejected between preemptive and dialyzed patients at baseline. It may suggest that patients' perception of HRQoL did not differ

between groups, but it could also be related to a lack of power. RS has been detected on mental HRQoL but not on physical HRQoL. We found similar uniform recalibration on the GH domain for preemptive and dialyzed kidney transplant patients only between baseline and 3-month after transplantation. Patients tended to report a better general health after KT than before given similar mental HRQoL. It could be that patients think they should feel better because KT is an opportunity, and they do not want to be ungrateful [19]. An additional or alternative explanation is that GH domain is composed of items such as "In general, how would you say your health is?" which are broadly worded items. Such general and unspecific items may leave room for interpretation and therefore may be more prone to RS detection [32, 33].

We found uniform recalibration on the VT domain, for preemptive patients only, between 3 (M3) and 6 (M6) months after transplantation. Preemptive patients tended to report a better perceived vitality at M6 compared to M3 given similar mental HRQoL. After transplantation, patients often experience restrictions and side effects of medication such as fatigue [19, 34], and the VT domain of the SF-36 questionnaire is often used as a marker of fatigue [35]. One explanation for this RS effect could be that preemptive patients probably expected to experience even greater levels of fatigue than they eventually did, which may not have been the case for the other patients already accustomed to the stress and fatigue associated with dialysis sessions [6, 7].

The means of latent variables with or without RS adjustment were very similar (Figures 3 and 4). The mental and physical HRQoL of patients tended to increase after KT in preemptive and dialyzed patients. Several studies assuming measurement invariance have shown better HRQoL (on SF-36 domains) in patients at 3 and 12 months after KT compared to before [6, 7]. A significant better physical HRQoL for preemptive compared to dialyzed patients was evidenced, which may suggest different supportive care needs between these patients before KT. The mental HRQoL mean was however similar at baseline between the two groups.

We observed small RS effects (Table 4), which may explain why small differences in estimated means were found whether RS was accounted for or not. In addition, a possible lack of power cannot be ruled out due to a small sample size compared to the number of estimated parameters in the SEM model. The proposed SEM procedure leads to a complex model with the estimation of many parameters. In our study, more than 100 parameters were estimated for the two-occasion SEM model for 374 included patients. Several authors have proposed recommendations regarding SEM and sample size requirements [36–39]. For example, Little [36] suggested a sample size such as 500, while Hays et al. [37] recommended that the number of cases should be 10 cases per estimated parameter. But, other authors recommend to be careful with these rules of thumb [38, 39]. However, we can highlight that our procedure allows to limit the number of estimated parameters by adding measurement occasions step-by-step.

Stringent approaches to model fit assessment are warranted but still debated, as for example the use of the χ^2 test of exact fit, due to its sensitivity to sample size and disputable inference testing aiming at not rejecting the null hypothesis [36, 40]. Alternative model fit indices have been developed (e.g. RMSEA, CFI, TLI, SRMR) but with arbitrary cut-off values to guide fit evaluation [36, 41] without providing evidence of local misfit. Hence, reporting of the χ^2 test and the former indices have been advised to be complemented by the inspection of standardized residual matrix [42], modification indices and associated expected parameter changes [43] in order to identify eventual misspecifications and local misfit. Our measurement model in step A was evaluated using the χ^2 test, which was significant, RMSEA and CFI, which indicated acceptable fit after addition of residual covariance that seems plausible according to the content of the SF36 domains. A further inspection of the model revealed that the majority of the residual covariance matrix components were <0.10 [42] and that modification indices were small with no sign of potential cross loadings. Moreover, no indication of misfit such as convergence

problems, inflated standard errors or negative variances appeared. Hence, this model was considered as an acceptable measurement model for the analysis given the data at hand.

A limitation of our study is the large number of tests performed in the procedure which may lead to chance findings. A word of caution is also warranted when lack of measurement invariance between groups or RS is detected on all domains. In these instances, it may not be wise to continue the procedure because the measurements are not comparable across the groups nor over time. Another limitation pertains to the assumption of linearity of measurement. Strategies for permitting arbitrary nonlinear relationships at the measurement level such as model-implied instrument variable-based SEM (MIIV-SEM) [44, 45] (REF) could be used and would be worth investigating.

It may be of interest for future research to include exogenous variables such as the type of transplant (deceased or living donor transplant) that may influence HRQoL and RS [19, 46]. Assessing changes in HRQoL and RS and between preemptive and dialyzed patients at the item level using item response or Rasch measurement theories [47, 48] could provide complementary insight to analyses performed at domain-level.

In conclusion, the procedure proposed in this paper allows detecting and taking into account RS on more than two occasions and between groups. In particular, our procedure also allows identifying the period where RS may have occurred. . Our study gives more insight into change in HRQoL after KT. Assessing measurement invariance can help targeting domains prone to more personalized psychological interventions according to previous dialysis experience.

References

1. Wolfe, R. A., Ashby, V. B., Milford, E. L., Ojo, A. O., Ettenger, R. E., Agodoa, L. Y., … Port, F. K. (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *The New England*

Journal of Medicine, *341*(23), 1725–1730. https://doi.org/10.1056/NEJM199912023412303

2. Tonelli, M., Wiebe, N., Knoll, G., Bello, A., Browne, S., Jadhav, D., … Gill, J. (2011). Systematic Review: Kidney Transplantation Compared with Dialysis in Clinically Relevant Outcomes. *American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons*, *11*(10), 2093–2109.

3. Jaar, B. G., Chang, A., & Plantinga, L. (2013). Can We Improve Quality of Life of Patients on Dialysis? *Clinical Journal of the American Society of Nephrology*, *8*(1), 1–4. https://doi.org/10.2215/CJN.11861112

4. Loos, C., Briançon, S., Frimat, L., Hanesse, B., & Kessler, M. (2003). Effect of End-Stage Renal Disease on the Quality of Life of Older Patients. *Journal of the American Geriatrics Society*, *51*(2), 229–233. https://doi.org/10.1046/j.1532-5415.2003.51062.x

5. Rebollo, P., Ortega, F., Baltar, J. M., Badía, X., Alvarez-Ude, F., Díaz-Corte, C., … Alvarez-Grande, J. (2000). Health related quality of life (HRQOL) of kidney transplanted patients: variables that influence it. *Clinical Transplantation*, *14*(3), 199–207.

6. Matsumura, S., Unagami, K., Okabe, S., Fukuda, H., Suzuki, A., Fuchinoue, S., … Ishida, H. (2018). Comparative Study on Variation of Quality of Life of Patients of Preemptive Kidney Transplantation and Nonpreemptive Kidney Transplantation. *Transplantation Proceedings*, *50*(10), 3321–3328. https://doi.org/10.1016/j.transproceed.2018.08.024

7. Mitsui, Y., Araki, M., Maruyama, Y., Yoshinaga, K., Sadahira, T., Wada, K., … Nasu, Y. (2020). Quality of Life and Mental Satisfaction Improve Slowly in Preemptive Kidney Transplantation Compared With Nonpreemptive Kidney Transplantation. *Transplantation Proceedings*. https://doi.org/10.1016/j.transproceed.2020.01.042

8. Bzoma, B., Walerzak, A., Dębska-Ślizień, A., Zadrożny, D., Śledziński, Z., & Rutkowski, B. (2016). Psychological Well-Being in Patients After Preemptive Kidney Transplantation. *Transplantation Proceedings*, *48*(5), 1515–1518. https://doi.org/10.1016/j.transproceed.2016.01.039

9. Mellenbergh, G. J. (1989). Item Bias and Item Response Theory. *International Journal of Educational Research*, *13*(2), 127–143.

10. Sprangers, M. A. G., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine*, *48*(11), 1507–1515. https://doi.org/10.1016/S0277-9536(99)00045-3

11. Schwartz, C. E., Bode, R., Repucci, N., Becker, J., Sprangers, M. A. G., & Fayers, P. M. (2006). The clinical significance of adaptation to changing health: a meta-analysis of response shift. *Quality of Life Research*, *15*(9), 1533–1550. https://doi.org/10.1007/s11136-006-0025-9

12. Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. https://doi.org/10.1177/014662169301700401

13. Sajobi, T. T., Brahmbatt, R., Lix, L. M., Zumbo, B. D., & Sawatzky, R. (2018). Scoping review of response shift methods: current reporting practices and recommendations. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *27*(5), 1133–1146. https://doi.org/10.1007/s11136-017-1751-x

14. Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*(3), 587–598. https://doi.org/10.1007/s11136-004-0830-y

15. Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11 Suppl 3), S78-94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

16. Lix, L. M., Chan, E. K. H., Sawatzky, R., Sajobi, T. T., Liu, J., Hopman, W., & Mayo, N. (2016). Response shift and disease activity in inflammatory bowel disease. *Quality of Life Research*, *25*(7), 1751–1760. https://doi.org/10.1007/s11136-015-1188-z

17. Verdam, M. G. E., & Oort, F. J. (2019). The Analysis of Multivariate Longitudinal Data: An Instructive Application of the Longitudinal Three-Mode Model. *Multivariate Behavioral Research*, *54*(4), 457–474. https://doi.org/10.1080/00273171.2018.1520072

18. Verdam, M. G. E., Oort, F. J., van der Linden, Y. M., & Sprangers, M. A. G. (2015). Taking into account the impact of attrition on the assessment of response shift and true change: a multigroup structural equation modeling approach. *Quality of Life Research*, *24*(3), 541–551. https://doi.org/10.1007/s11136-014-0829-y

19. Schipper, K., Abma, T. A., Koops, C., Bakker, I., Sanderman, R., & Schroevers, M. J. (2014). Sweet and sour after renal transplantation: A qualitative study about the positive and negative consequences of renal transplantation. *British Journal of Health Psychology*, *19*(3), 580–591. https://doi.org/10.1111/bjhp.12057

20. Smith, D., Loewenstein, G., Jepson, C., Jankovich, A., Feldman, H., & Ubel, P. (2008). Mispredicting and misremembering: patients with renal failure overestimate improvements in quality of life after a kidney transplant. *Health Psychology*, *27*(5), 653–658. https://doi.org/10.1037/a0012647

21. Sébille, V., Hardouin, J.-B., Giral, M., Bonnaud-Antignac, A., Tessier, P., Papuchon, E., … Meurette, A. (2016). Prospective, multicenter, controlled study of quality of life, psychological adjustment process and medical outcomes of patients receiving a preemptive kidney transplant compared to a similar population of recipients after a dialysis period of less than three years – The PreKit-QoL study protocol. *BMC Nephrology*, *17*. https://doi.org/10.1186/s12882-016-0225-7

22. Leplège, A., Ecosse, E., Pouchot, J., Coste, J., & Perneger, T. (2011). *Le questionnaire MOS SF-36 : Manuel d'utilisateur et guide d'interprétation des scores*. Paris: Estem.

23. Oort, F. J. (2005). Towards a Formal Definition of Response Shift (In Reply to G.W. Donaldson). *Quality of Life Research*, *14*(10), 2353–2355. https://doi.org/10.1007/s11136-005-3978-1

24. Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, *62*(11), 1126–1137. https://doi.org/10.1016/j.jclinepi.2009.03.013

25. Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

26. Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

27. Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258.

28. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255.

29. Vigneau, E., & Qannari, E. M. (2003). Clustering of Variables Around Latent Components. *Communications in Statistics - Simulation and Computation*, *32*(4), 1131–1150. https://doi.org/10.1081/SAC-120023882

30. Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New-York: Wiley.

31. Verdam, M. G. E., Oort, F. J., & Sprangers, M. A. G. (2017). Structural equation modeling–based effect-size indices were used to evaluate and interpret the impact of response shift effects. *Journal of Clinical Epidemiology*, *85*, 37–44. https://doi.org/10.1016/j.jclinepi.2017.02.012

32. King-Kallimanis, B. L., Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Structural equation modeling of health-related quality-of-life data illustrates the measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology*, *62*(11), 1157–1164. https://doi.org/10.1016/j.jclinepi.2009.04.004

33. Vanier, A., Leplège, A., Hardouin, J.-B., Sébille, V., & Falissard, B. (2015). Semantic primes theory may be helpful in designing questionnaires such as to prevent response shift. *Journal of Clinical Epidemiology*, *68*(6), 646–654. https://doi.org/10.1016/j.jclinepi.2015.01.023

34. Rodrigue, J. R., Mandelbrot, D. A., Hanto, D. W., Johnson, S. R., Karp, S. J., & Pavlakis, M. (2011). A cross-sectional study of fatigue and sleep quality before and after kidney transplantation. *Clinical Transplantation*, *25*(1), E13-21. https://doi.org/10.1111/j.1399-0012.2010.01326.x

35. Jhamb, M., Jhamb, M., Liang, K., Yabes, J., Steel, J. L., Dew, M. A., … Unruh, M. (2013). Prevalence and Correlates of Fatigue in Chronic Kidney Disease and End-Stage Renal Disease: Are Sleep Disorders a Key to Understanding Fatigue? *American Journal of Nephrology*, *38*(6), 489–495. https://doi.org/10.1159/000356939

36. Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY, US: Guilford Press.

37. Hays, R. D., Revicki, D., & Coyne, K. S. (2005). Application of structural equation modeling to health outcomes research. *Evaluation & the Health Professions*, *28*(3), 295–309. https://doi.org/10.1177/0163278705278277

38. Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and Psychological Measurement*, *76*(6), 913–934. https://doi.org/10.1177/0013164413495237

39. Wang, Y. A., & Rhemtulla, M. (2021). Power Analysis for Parameter Estimation in Structural Equation Modeling: A Discussion and Tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–17.

40. Hayduk, L. A. (2014). Shame for disrespecting evidence: the personal consequences of insufficient respect for structural equation model testing. *BMC Medical Research Methodology*, *14*(1), 124. https://doi.org/10.1186/1471-2288-14-124

41. Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, *6*, 1715. https://doi.org/10.3389/fpsyg.2015.01715

42. Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (Fourth Edition.). New-York, NY: Guilford Press.

43. Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, *16*(4), 561–582. https://doi.org/10.1080/10705510903203433

44. Bollen, K. A. (2019). Model Implied Instrumental Variables (MIIVs): An Alternative Orientation to Structural Equation Modeling. *Multivariate Behavioral Research*, *54*(1), 31–46. https://doi.org/10.1080/00273171.2018.1483224

45. Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2020). Latent variable GIMME using model implied instrumental variables (MIIVs). *Psychological Methods*, *25*(2), 227–242. https://doi.org/10.1037/met0000229

46. Gill, P., & Lowes, L. (2008). Gift exchange and organ donation: donor and recipient experiences of live related kidney transplantation. *International Journal of Nursing Studies*, *45*(11), 1607–1617. https://doi.org/10.1016/j.ijnurstu.2008.03.004

47. Guilleux, A., Blanchin, M., Vanier, A., Guillemin, F., Falissard, B., Schwartz, C. E., … Sébille, V. (2015). RespOnse Shift ALgorithm in Item response theory (ROSALI) for response shift detection with missing data in longitudinal patient-reported outcome studies. *Quality of Life Research*, *24*(3), 553–564. https://doi.org/10.1007/s11136-014-0876-4

48. Blanchin, M., Guilleux, A., Hardouin, J.-B., & Sébille, V. (2020). Comparison of structural equation modelling, item response theory and Rasch measurement theory-based methods for response shift detection at item level: A simulation study. *Statistical Methods in Medical Research*, *29*(4), 1015–1029. https://doi.org/10.1177/0962280219884574

Tables

Table 1: Characteristics of the preemptive and dialyzed patients at baseline.

	Preemptive group (n = 178)	Dialysis group (n = 196)	p-value
Center, n (%)			0.12
Lyon	15 (8 %)	27 (14 %)	
Nantes	124 (70 %)	118 (60 %)	
Nice	39 (22 %)	51 (26 %)	
Sociodemographic data			
Age (years, mean ± SD [range])	57 ± 14 [20-83]	55 ± 15 [21-86]	0.41
Gender, n (%)			0.81
Female	63 (35 %)	67 (34 %)	
Male	115 (65 %)	129 (66%)	
Transplantation associated data			
Types of transplants, n (%)			0.08
Deceased donor transplant	109 (61 %)	137 (70 %)	
Living donor transplant	69 (39 %)	59 (30 %)	
Initial disease, n (%)			0.18
Diabetes	11 (7 %)	25 (14 %)	
Glomerulonephritis	32 (20 %)	34 (18 %)	
Tubulo-interstitial nephritis	19 (12 %)	14 (8 %)	
Vascular disease	18 (11 %)	26 (14 %)	
Others	79 (50 %)	84 (46 %)	
Diabetes, n (%)	24 (7 %)	46 (13 %)	0.02
Cardiovascular, n (%)	149 (41 %)	171 (47 %)	0.26
Cancer, n (%)	22 (6 %)	19 (5 %)	0.41
Time on waiting list transplantation (months, median (IQR))	10 (15)	13 (17)	0.24
Time on dialysis (months, mean ± SD [range])	n/a	13.5 ± 9.1 [0-35]	n/a

SD standard deviation, IQR interquartile range, n/a not applicable

Table 2: Overall goodness-of-fit and differences between model fit in the procedure of investigation of measurement invariance

Model	Description	χ^2	df	χ^2/df	RMSEA	CFI	LR	p-value
Detection of measurement non-invariance between groups at baseline								
Step A	Non-invariance between groups	66.51	31	2.15	0.078	0.972	-	-
Step B	Invariance between groups	93.91	50	1.88	0.069	0.966	27.40	0.096
Response shift detection between baseline and 3-month visit after transplantation								
Step 1	Response shift model	340.72	188	1.81	0.066	0.949	-	-
Step 2	No response shift model	451.35	232	1.95	0.071	0.927	110.64	< 0.001
Step 3	Final model	411.47	231	1.78	0.065	0.940	39.89*	< 0.001
Response shift detection between 3- and 6-month visits after transplantation								
Step 1	Response shift model	971.78	477	2.04	0.074	0.907	-	-
Step 2	No response shift model	1033.68	521	1.98	0.073	0.904	61.90	0.039
Step 3	Final model	1021.74	520	1.96	0.072	0.906	11.94*	< 0.001

df, degree of freedom; RMSEA, Root-Mean-Square Error of Approximation; CFI, Comparative Fit Index; LR, likelihood ratio statistic; *: Likelihood Ratio (LR) test between models of the step 2 and 3; -: not applicable

Table 3: Parameters estimates for the final model of the procedure for three measurement occasions

	Physical HRQoL				Mental HRQoL			
	PF	RP	BP	RE	SF	VT	MH	GH
Intercepts for 3 measurement occasions								
Dialyzed group	0.701	0.359	0.614	0.488	0.663	0.451	0.644	0.463/0.517/0.517
Preemptive group	0.701	0.359	0.614	0.488	0.663	0.451/0.451/0.485	0.644	0.463/0.517/0.517
Residual variances for 3 measurement occasions								
Dialyzed group	0.035	0.068	0.030	0.095	0.023	0.009	0.015	0.020
Preemptive group	0.035	0.068	0.030	0.095	0.023	0.009	0.015	0.020
Factor loadings for 3 measurement occasions								
Dialyzed group	0.120	0.293	0.144	0.273	0.152	0.166	0.113	0.107
Preemptive group	0.120	0.293	0.144	0.273	0.152	0.166	0.113	0.107
	Physical HRQoL				Mental HRQoL			
	Before transplantation	3 months after transplantation		6 months after transplantation	Before transplantation	3 months after transplantation		6 months after transplantation
Common factor variances								
Dialyzed group	1	1	1	1	1	1	1	1
Preemptive group	0.926	0.926	0.926	0.926	1.010	1.010	1.010	1.010
Common factor means								
Dialyzed group	0	0.196	0.196	0.444	0	0.610	0.610	0.685
Preemptive group	0.326	0.296	0.296	0.662	-0.002	0.630	0.630	0.635

Notes: Parameters estimates separated by a slash represent estimates for each measurement occasion when parameters are not constrained over time. For computational convenience, scores of the SF-36 questionnaire used for the SEM procedure have been divided by 100. Abbreviations: PF, Physical Functioning; RP, Role Physical; BP, Bodily Pain; RE, Role Emotional; SF, Social Functioning; VT, Vitality; MH, Mental Health; GH, General Health. Parameters estimates for common factor correlation between latent variables are presented in Supplementary Table B.

Table 4: Effect-sizes of observed change, response shift contribution, and target change for the final model of the procedure in preemptive and dialyzed groups

	Preemptive group				Dialyzed group			
	Type of response shift	Observed change	Response shift contribution	Target change contribution	Type of response shift	Observed change	Response shift contribution	Target change contribution
Between baseline and 3 months after transplantation								
Physical Functioning (PF)		-0.02		-0.02		0.10		0.10
Role Physical (RP)		-0.02		-0.02		0.12		0.12
Bodily Pain (BP)		-0.02		-0.02		0.11		0.11
Role Emotional (RE)		-0.02		-0.02		0.11		0.11
Social Functioning (SF)		0.36		0.36		0.36		0.36
Vitality (VT)		0.45		0.45		0.44		0.44
Mental Health (MH)		0.38		0.38		0.40		0.40
General Health (GH)	Uniform recalibration	0.59	0.26	0.33	Uniform recalibration	0.56	0.25	0.31
Between 3 months and 6 months after transplantation								
Physical Functioning (PF)		0.23		0.23		0.17		0.17
Role Physical (RP)		0.22		0.22		0.17		0.17
Bodily Pain (BP)		0.21		0.21		0.16		0.16
Role Emotional (RE)		0.21		0.21		0.14		0.14
Social Functioning (SF)		< 0.01		< 0.01		0.05		0.05
Vitality (VT)	Uniform recalibration	0.18	0.17	< 0.01		0.06		0.06
Mental Health (MH)		< 0.01		< 0.01		0.05		0.05
General Health (GH)		< 0.01		< 0.01		0.04		0.04

Effect-sizes values of 0.2, 0.5, and 0.8 are considered "small", "medium", and "large"

Table 5: Effect estimates and standard error for group effect, time effect and their interaction using Wald tests, for the latent variables Mental and Physical HRQoL with or without RS adjustment

		Group effect at baseline		Time effect between baseline and 3-months visit after kidney transplantation		Time effect between 3- and 6-months visit after kidney transplantation		Interaction between group and time (baseline and 3-months visit after kidney transplantation)		Interaction between group and time (3- and 6-months visit after kidney transplantation)	
		Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
With RS adjustment											
Mental HRQoL	Dialyzed group	-0.002 (0.11)	0.98	0.61 (0.10)	< 0.001	0.07 (0.08)	0.38	0.02 (0.14)	0.87	-0.07 (0.13)	0.60
	Preemptive group			0.63 (0.10)	< 0.001	0.005 (0.10)	0.96				
Physical HRQoL	Dialyzed group	0.33 (0.12)	0.006	0.20 (0.10)	0.043	0.25 (0.09)	0.004	-0.23 (0.15)	0.12	0.12 (0.13)	0.38
	Preemptive group			-0.003 (0.11)	0.78	0.37 (0.10)	< 0.001				
Without RS adjustment											
Mental HRQoL	Dialyzed group	-0.02 (0.11)	0.84	0.65 (0.10)	< 0.001	0.064 (0.08)	0.45	0.03 (0.14)	0.82	0.05 (0.12)	0.66
	Preemptive group			0.69 (0.11)	< 0.001	0.12 (0.09)	0.20				
Physical HRQoL	Dialyzed group	0.33 (0.12)	0.006	0.20 (0.10)	0.041	0.25 (0.09)	0.004	0.23 (0.15)	0.12	0.12 (0.13)	0.38
	Preemptive group			-0.03 (0.11)	0.78	0.36 (0.10)	< 0.001				

The dialyzed group is the reference group. SE, standard error

Figure legends

Figure 1: Procedure to detect, assess and account for lack of measurement invariance between groups and response shift with more than two occasions.
Notes: *t* corresponds to the number of measurement occasions with *t*=1 for the baseline visit before kidney transplantation, *t*=2 for the 3-month visit after transplantation, and *t*=3 for the 6-month visit after transplantation.

Figure 2: Measurement model at baseline (last visit before the kidney transplantation) for the SF-36 questionnaire (model for step A of the Part 1 of the procedure).
Notes: The SEM model structure is the same for the dialyzed and preemptive patients group. To simplify the figure, the SEM structure is represented for one group only. Circles represent latent variables and rectangles represent observed variables. Single-headed arrows represent paths. Double-headed arrows represent covariation. Abbreviation: PF, Physical Functioning; RP, Role Physical; BP, Bodily Pain; RE, Role Emotional; SF, Social Functioning; VT, Vitality; MH, Mental Health; GH, General Health; Res, residual variance.

Figure 3: Change in the mean of the latent variable physical HRQoL by group across measurement occasions.
Notes: Circles and triangles symbol represent the dialyzed and preemptive patients group, respectively. The solid lines represent the means of the latent variable taking into account response shift. The dash lines represent the means of the latent variable without taking into account response shift; Measurement occasions are the last visit before kidney transplantation (baseline), 3- and 6-months visit after transplantation; *: significant time effect.

Figure 4: Change in the mean of the latent variable mental HRQoL by group across measurement occasions.
Notes: Circles and triangles symbol represent the dialyzed and preemptive patients group, respectively. The solid lines represent the means of the latent variable taking into account response shift. The dash lines represent the means of the latent variable without taking into account response shift; Measurement occasions are the last visit before kidney transplantation (baseline), 3- and 6-months visit after transplantation; *: significant time effect.

Figure 1

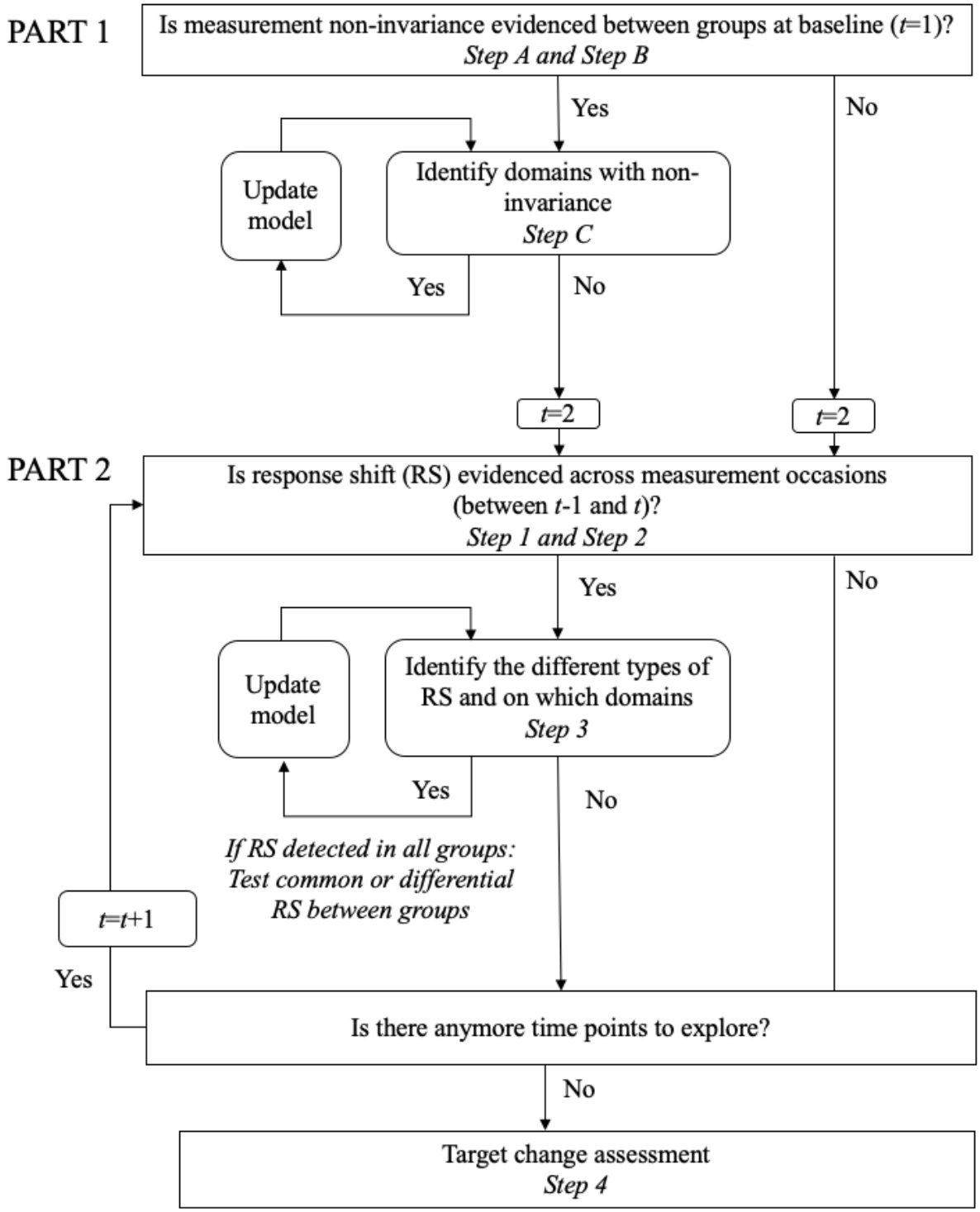

PART 1
Is measurement non-invariance evidenced between groups at baseline (t=1)?
Step A and Step B
Yes
No
Update model
Identify domains with non-invariance
Step C
Yes
No
t=2
t=2
PART 2
Is response shift (RS) evidenced across measurement occasions (between t-1 and t)?
Step 1 and Step 2
Yes
No
Update model
Identify the different types of RS and on which domains
Step 3
Yes
No
If RS detected in all groups:
Test common or differential
RS between groups
t=t+1
Yes
Is there anymore time points to explore?
No
Target change assessment
Step 4

Figure 2

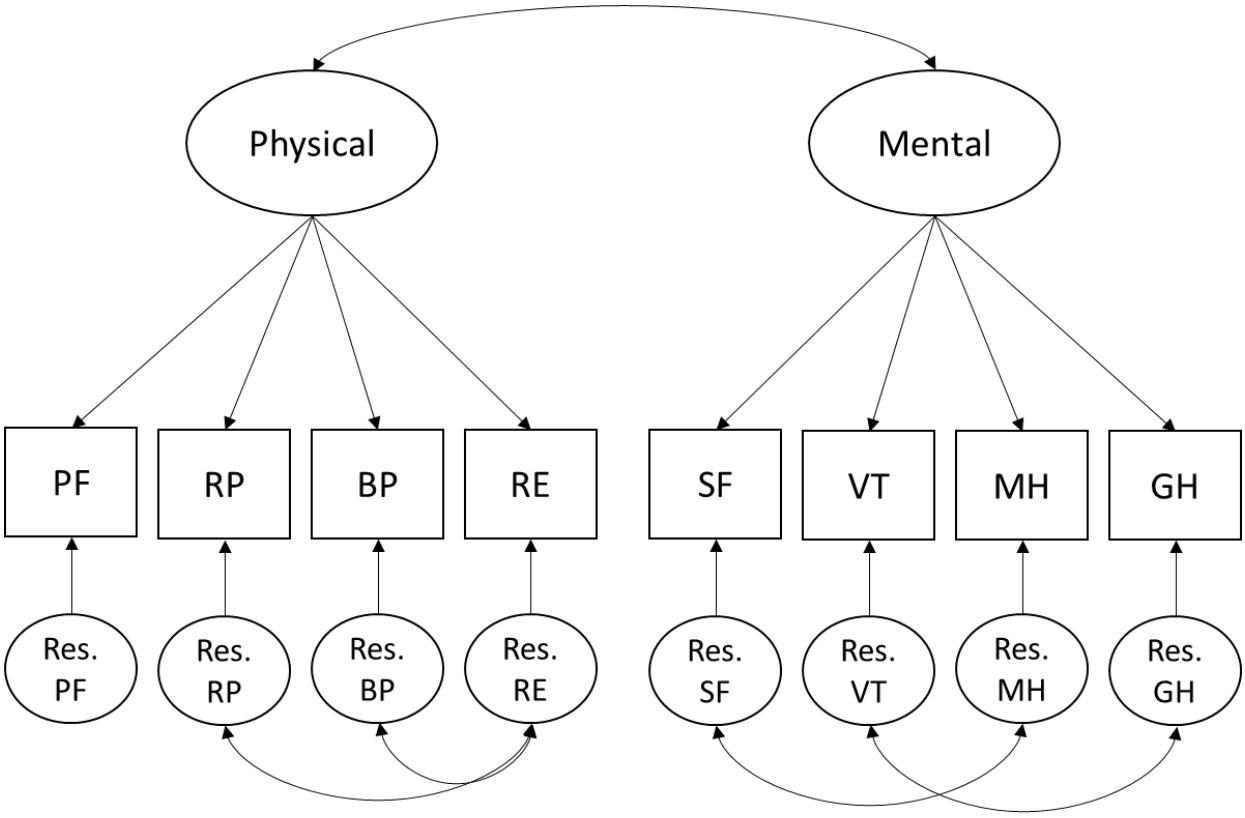

Physical
Mental
PF
RP
BP
RE
SF
VT
MH
GH
Res. PF
Res. RP
Res. BP
Res. RE
Res. SF
Res. VT
Res. MH
Res. GH

Figure 3

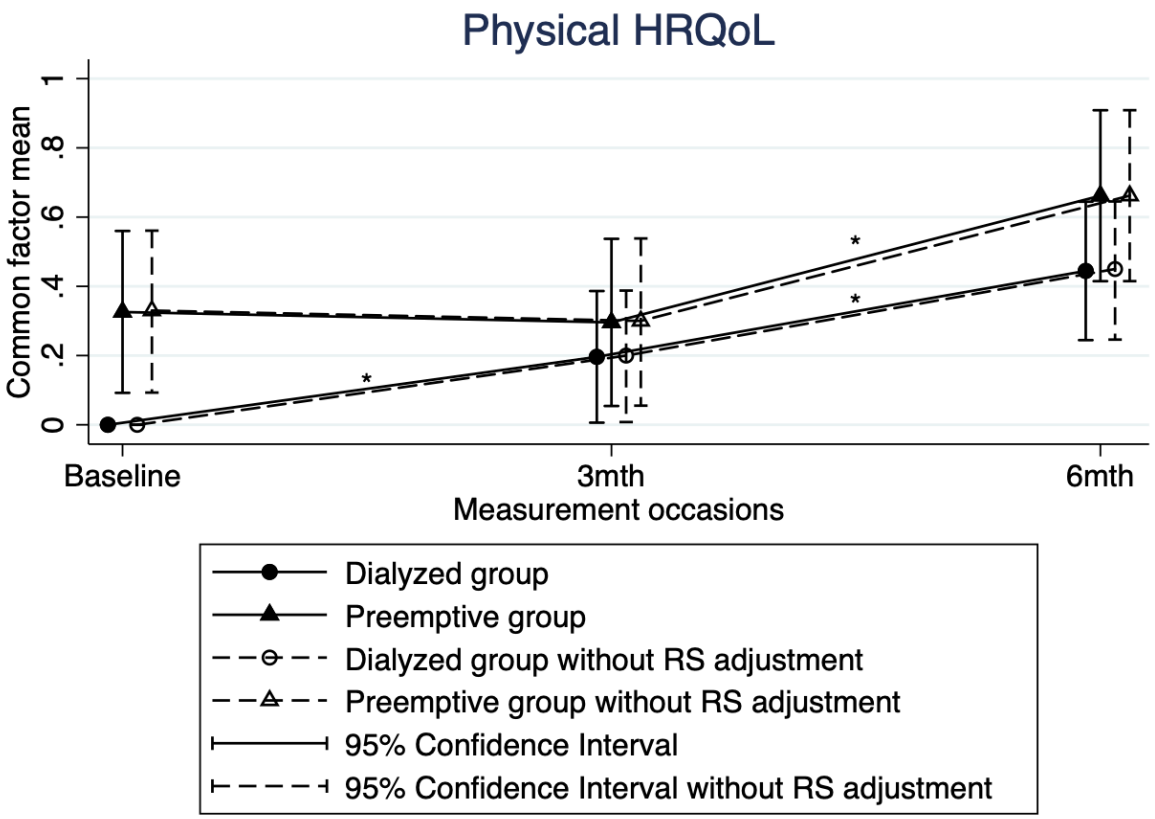
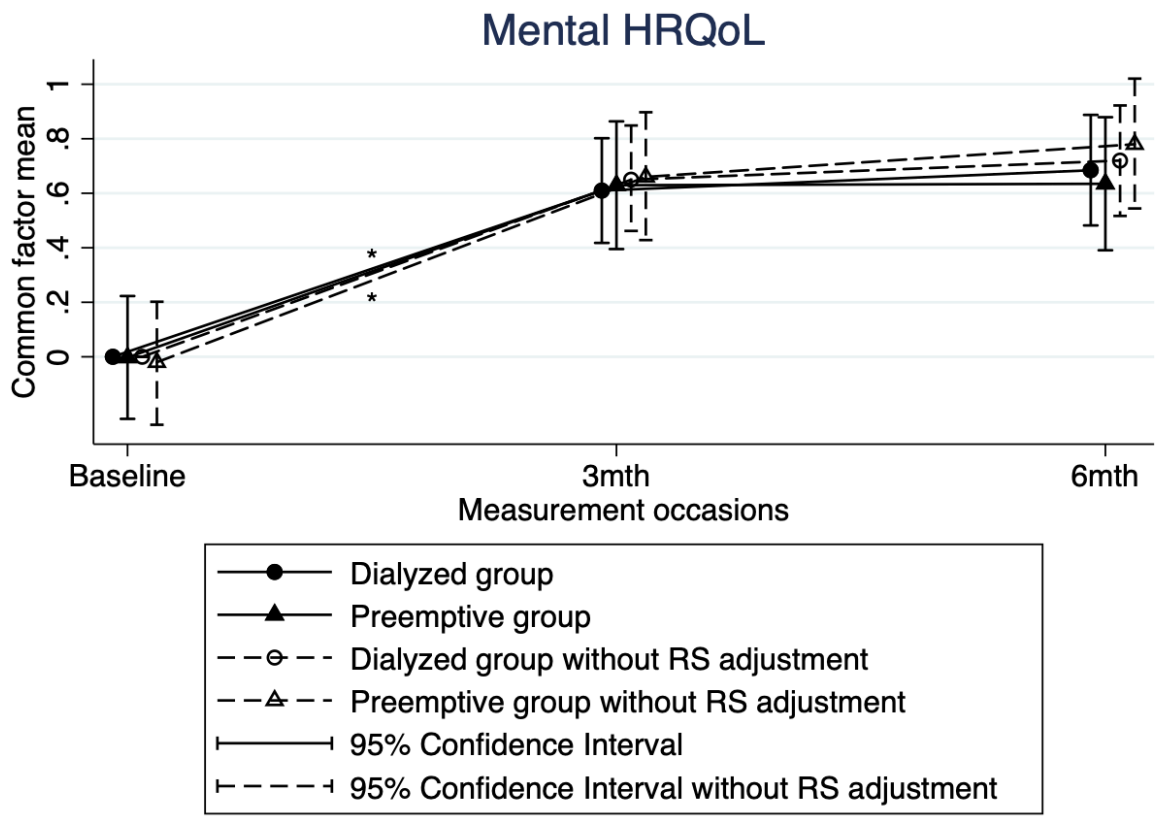

Physical HRQoL
Common factor mean
0
.2
.4
.6
.8
1
Baseline
3mth
6mth
Measurement occasions
*
*
*
Dialyzed group
Preemptive group
Dialyzed group without RS adjustment
Preemptive group without RS adjustment
95% Confidence Interval
95% Confidence Interval without RS adjustment

Figure 4


Mental HRQoL
Common factor mean
0
.2
.4
.6
.8
1
Baseline
3mth
6mth
Measurement occasions
Dialyzed group
Preemptive group
Dialyzed group without RS adjustment
Preemptive group without RS adjustment
95% Confidence Interval
95% Confidence Interval without RS adjustment

Supplementary Table A: Means, standard deviations and estimates of skewness and kurtosis for each domains of the SF-36 questionnaire, in preemptive and dialyzed group, at each measurement occasions

	Before transplantation				3 months after transplantation				6 months after transplantation			
Domains	Missing data	Mean (SD)	Skewness	Kurtosis	Missing data	Mean (SD)	Skewness	Kurtosis	Missing data	Mean (SD)	Skewness	Kurtosis
Dialyzed group												
PF	2	69.60 (24.03)	-0.85	2.90	11	72.56 (23.77)	-0.99	3.36	22	76.17 (22.41)	-1.13	3.86
RP	4	36.81 (38.79)	0.57	1.79	13	40.12 (39.97)	0.40	1.58	23	51.73 (44.15)	-0.07	1.25
BP	7	59.76 (22.51)	0.004	2.49	17	67.95 (23.52)	-0.28	2.30	28	67.40 (23.95)	-0.30	2.27
RE	5	50.09 (44.70)	0.01	1.22	13	51.09 (42.71)	-0.02	1.32	22	63.70 (42.55)	-0.55	1.57
SF	10	66.26 (22.46)	0.02	2.03	25	73.76 (21.93)	-0.67	2.77	25	76.68 (20.56)	-0.41	2.09
VT	5	45.34 (19.10)	0.19	2.61	13	55.60 (19.70)	-0.19	2.64	23	57.22 (20.54)	-0.17	2.62
MH	2	63.63 (19.09)	-0.42	2.52	13	70.77 (17.87)	-0.73	3.16	23	71.21 (16.80)	-0.54	2.79
GH	1	46.44 (18.45)	0.24	2.18	13	59.02 (19.12)	-0.10	2.73	21	57.23 (19.91)	-0.002	2.57
Preemptive group												
PF	3	76.05 (22.03)	-1.19	4.07	28	75.11 (23.52)	-1.33	4.29	30	77.61 (24.31)	-1.54	4.87
RP	2	46.54 (39.37)	0.13	1.47	28	42.39 (41.75)	0.33	1.46	29	57.05 (39.21)	-0.32	1.57
BP	4	63.81 (22.80)	-0.21	2.44	32	68.25 (22.52)	-0.28	2.38	34	70.76 (23.29)	-0.66	2.81
RE	2	58.81 (41.67)	-0.37	1.49	26	56.14 (43.47)	-0.23	1.32	28	69.11 (39.78)	-0.85	2.10
SF	3	68.29 (23.89)	-0.44	2.66	26	74.42 (23.17)	-0.80	3.08	30	75.93 (23.42)	0.85	2.91
VT	2	43.88 (20.53)	0.16	2.64	25	55.91 (20.43)	-0.39	2.97	27	58.58 (18.84)	-0.44	2.72
MH	1	65.85 (17.07)	-0.45	2.85	25	71.72 (17.65)	-0.71	3.13	26	72.09 (17.50)	-0.80	3.43
GH	3	45.24 (17.92)	0.21	2.81	24	59.20 (18.21)	-0.22	2.73	27	58.64 (18.02)	-0.21	2.45

SD, Standard Deviation; Abbreviations: PF, Physical Functioning; RP, Role Physical; BP, Bodily Pain; RE, Role Emotional; SF, Social Functioning; VT, Vitality; MH, Mental Health; GH, General Health.

Supplementary Table B: Parameter estimates for correlations between latent variables for the final model of the procedure for three measurement occasions

		Physical HRQoL			Mental HRQoL		
		Before transplantation	3 months after transplantation	6 months after transplantation	Before transplantation	3 months after transplantation	6 months after transplantation
Dialyzed group							
Physical HRQoL	Before transplantation	1					
	3 months after transplantation	0.439	1		0		
	6 months after transplantation	0.392	0.576	1	0	0	
Mental HRQoL	Before transplantation	0.755			1		
	3 months after transplantation	0	0.414		0.301	1	
	6 months after transplantation	0	0	0.530	0.229	0.455	1
Preemptive group							
Physical HRQoL	Before transplantation	0.926					
	3 months after transplantation	0.168	0.926		0		
	6 months after transplantation	0.283	0.336	0.926	0	0	
Mental HRQoL	Before transplantation	0.799			1.010		
	3 months after transplantation	0	0.543		0.306	1.010	
	6 months after transplantation	0	0	0.624	0.165	0.426	1.010

Notes: Parameters estimates separated by a slash represent estimates for each measurement occasion when parameters are not constrained over time. For computational convenience, scores of the SF-36 questionnaire used for the SEM procedure have been divided by 100. Abbreviations: PF, Physical Functioning; RP, Role Physical; BP, Bodily Pain; RE, Role Emotional; SF, Social Functioning; VT, Vitality; MH, Mental Health; GH, General Health.