The Knowledge Assessment of Renal Transplantation (KART) 2.0: Development and Validation of Chronic Kidney Disease and Transplant Knowledge Scales

Supplemental Material

Contents

Rationale for Choice of Topics to Include in the KART 2.0 Measure

The primary rationale for choice of topics for the KART 2.0 items was to capture content that could be featured in educational interventions and key counselling sessions with patients in clinic. One investigator (ADW) and her team initially started identifying topics in formative qualitative work conducted in 2006-2008.[1,2] Since that time, during the course of developing, launching, and testing various educational interventions, these topics were revisited less formally through a patient panel convened specifically to give feedback about these interventions. Input from those panels suggested that more content on risks of living donation and risks of receiving a transplant. In addition, in developing the partnership with Kaiser Permanente Southern California, a need for educational content addressing basic facts about CKD. Items written to address topics uncovered by ADW's formative work were featured in KART 1.0 (17 of the starting items for KART 2.0). To find items that cover these additional topics (risks of living donation, risks of transplant receipt, basic facts about CKD), we conducted an informal literature review; our goal was to find as many extant items as possible to reduce the number of new items that needed to be written. During this process, we identified a previously-published knowledge scale developed by another investigator (KLC) that included items on basic facts about CKD (11 of the 37 starting items). Our literature search did not locate any suitable items to include around risks of living donation and transplant, so new items were written to cover these topics.

Supplemental Table 1. Frequency and Proportion of Raw Responses to Each Item (N=977)

Item	N (%)
Does the kidney make urine? (CKD1)	
Yes	624 (64%)
No	160 (16%)
Don't Know	184 (19%)
Missing	9 (1%)
Does the kidney make clean blood? (CKD2)	
Yes	744 (76%)
No	75 (8%)
Don't Know	148 (15%)
Missing	10 (1%)
Does the kidney help keep bones healthy? (CKD3)	
Yes	510 (52%)
No	120 (12%)
Don't Know	338 (35%)
Missing	9 (1%)
Does the kidney help keep blood pressure normal? (CKD4)	
Yes	655 (67%)
No	83 (9%)
Don't Know	229 (23%)
Missing	10 (1%)
Shortness of breath is a symptom of chronic kidney disease. (CKD5)	
True	437 (45%)
False	158 (16%)
Don't Know	374 (38%)
Missing	8 (1%)
Confusion is a symptom of chronic kidney disease. (CKD6)	
True	418 (43%)
False	139 (14%)
Don't Know	409 (42%)
Missing	11 (1%)
Increased fatigue is a symptom of chronic kidney disease. (CKD7)	
True	756 (77%)
False	38 (4%)
Don't Know	174 (18%)
Missing	9 (1%)
Chronic kidney disease increases a person's chance for a heart attack. (CKD8)	
True	601 (62%)
False	56 (6%)
Don't Know	313 (32%)
Missing	7 (<1%)
Unusual itching is a symptom of chronic kidney disease. (CKD9)	
True	458 (47%)
False	99 (10%)
Don't Know	409 (42%)
Missing	11 (1%)
Nausea and/or vomiting is a symptom of chronic kidney disease. (CKD10)	
True	492 (50%)
False	114 (12%)
Don't Know	360 (37%)
Missing	11 (1%)
Metal taste/bad taste in the mouth is a symptom of chronic kidney disease. (CKD11)	
True	451 (46%)
False	106 (11%)

Don’t Know	405 (42%)
Missing	15 (1%)
Patients older than 70 years can receive transplants. (KTX1)	
True	294 (30%)
False	162 (17%)
Don’t Know	510 (52%)
Missing	11 (1%)
The transplant team will let a living donor back out from donating on the day of the surgery. (KTX2)	
True	489 (50%)
False	57 (6%)
Don’t Know	416 (43%)
Missing	15 (1%)
Patients have better health if they receive a transplant before starting dialysis. (KTX3)	
True	390 (40%)
False	89 (9%)
Don’t Know	490 (50%)
Missing	8 (1%)
Compared to having two working kidneys, dialysis works just as well as getting a transplant. (KTX4)	
True	207 (21%)
False	344 (35%)
Don’t Know	416 (43%)
Missing	10 (1%)
Individuals who donate a kidney have a slightly higher chance of high blood pressure in the future compared to people who don’t donate. (KTX5)	
True	160 (16%)
False	92 (9%)
Don’t Know	713 (73%)
Missing	12 (1%)
The transplant team cannot tell the kidney patient anything about the donor unless the donor agrees. (KTX6)	
True	529 (54%)
False	31 (3%)
Don’t Know	405 (41%)
Missing	12 (1%)
You can slow down how fast your kidneys fail. (KTX7)	
True	608 (62%)
False	139 (14%)
Don’t Know	220 (22%)
Missing	10 (1%)
In general, a living kidney donor will need to stay in the hospital more than 1 week after the surgery. (KTX8)	
True	199 (20%)
False	158 (16%)
Don’t Know	612 (63%)
Missing	8 (1%)
All living donors transportation costs to and from the transplant center are covered by Medicare and/or private insurance. (KTX9)	
True	229 (23%)
False	99 (10%)
Don’t Know	638 (65%)
Missing	11 (1%)
In general, patients can live longer with a kidney transplant than if they stayed on dialysis. (KTX10)	
True	555 (57%)
False	59 (6%)
Don’t Know	353 (36%)
Missing	10 (1%)

After a patient is listed, they don't need to return to the transplant center again until a matching kidney is found. (KTX11)	
True	165 (17%)
False	274 (28%)
Don't Know	527 (54%)
Missing	11 (1%)
Donors usually feel the most pain the first week after surgery. (KTX12)	
True	309 (32%)
False	21 (2%)
Don't Know	634 (65%)
Missing	13 (1%)
In general, most people on dialysis are happier with the quality of their lives than people with transplants. (KTX13)	
True	54 (6%)
False	480 (49%)
Don't Know	433 (44%)
Missing	10 (1%)
Transplant recipients are at risk of developing skin cancer. (KTX14)	
True	115 (12%)
False	133 (14%)
Don't Know	719 (74%)
Missing	10 (1%)
Kidney transplants that occur before a patient starts dialysis generally last longer than other transplants. (KTX15)	
True	185 (19%)
False	71 (7%)
Don't Know	703 (72%)
Missing	18 (2%)
Tremors are a possible side effect of anti-rejection medications. (KTX16)	
True	260 (27%)
False	35 (4%)
Don't Know	668 (68%)
Missing	14 (1%)
If a patient waits long enough on the wait list, a matching kidney from someone who has died will definitely become available. (KTX17)	
True	284 (29%)
False	277 (28%)
Don't Know	405 (42%)
Missing	11 (1%)
Compared to transplants from donors who have died, how long do transplants from living donors last? (KTX18)	
A shorter amount of time	30 (3%)
A longer amount of time	178 (18%)
The Same Amount of Time	117 (12%)
Don't Know	646 (66%)
Missing	6 (1%)
How long do patients usually wait on the waiting list for a kidney from someone who has died? (KTX19)	
Less than 1 year	27 (3%)
1-2 years	32 (3%)
3-5 years	102 (10%)
More than 5 years	307 (31%)
Don't know	505 (52%)
Missing	4 (1%)
What is the chance that a donor would die while undergoing transplant surgery? (KTX20)	
Less than 1%	182 (19%)

3%	51 (5%)
10%	28 (3%)
25%	23 (2%)
Don't Know	685 (70%)
Missing	8 (1%)
In general, after surgery, how long does it take for most donors to return to their normal daily activities? (KTX21)	
1 week	85 (8%)
1 month	144 (15%)
3 months	94 (9%)
6 months	57 (6%)
Don't know	592 (61%)
Missing	5 (1%)
Do donors have to pay for testing and hospitalization related to kidney donation? (KTX22)	
Yes	57 (6%)
No	285 (29%)
Don't Know	627 (64%)
Missing	8 (1%)
After a transplant, how long does the US Government pay for most of the costs of transplant medications? (KTX23)	
1 year	37 (4%)
3 years	36 (4%)
10 years	4 (<1%)
For the rest of the recipient's life	86 (9%)
Don't know	807 (82%)
Missing	7 (1%)
How long have doctors been doing transplants using living donors? (KTX24)	
2 years	8 (1%)
10 years	32 (3%)
25 years	135 (14%)
Over 50 years	120 (12%)
Don't Know	679 (70%)
Missing	3 (<1%)
About what percentage of all transplanted kidneys keep working for at least 1 year? (KTX25)	
50%	68 (7%)
75%	112 (11%)
90%	133 (14%)
Don't Know	656 (67%)
Missing	8 (1%)
What is the chance that a recipient would die while undergoing transplant surgery? (KTX26)	
Less than 1%	98 (10%)
3%	76 (8%)
10%	39 (4%)
25%	42 (4%)
Don't Know	714 (73%)
Missing	8 (1%)

Detailed Item Response Theory Methods

The two item properties focused on by the item response theory model used in this paper – the two-parameter logistic model (2PL) – include the difficulty threshold (*b*), defined as the location on the knowledge continuum where there is a 50% probability of answering below vs. above the threshold. *b* is presented in the z-metric with population mean = 0 and standard deviation of 1. For example, an item with *b* = 1 has a difficulty such that respondents with transplant knowledge of 1 standard deviation above the population mean have a 50% probability of giving a correct response. The second item property, the discrimination parameter (*a*), is an estimate of how well items differentiate between patients of lower and higher levels of knowledge, with higher values indicating better discrimination. *a* is also on the z-metric. The discrimination parameter can also be presented in the more familiar factor loading metric. These properties can be visualized with item characteristic curves that plot the latent trait estimate (theta) on the x axis and the probability of response on the y axis. In addition, the test's performance can be visualized using a plot of its information (for each item and for the total test), which is equivalent to reliability. Instead of a single reliability estimate for the overall test, item response theory provides a way to determine reliability for patients with different levels of knowledge, which in turn is useful to determining which patients the measure works best for. The 2PL model has the assumptions of unidimensionality, monotincity, and local dependence. Details about tests of these assumptions are given below.

We fit 2PL models using the MIRT package in R.[3] We fit two separate models, one for transplant knowledge and one for CKD knowledge. We used the results of these models to refine the item set. We plotted item and test information functions to examine reliability to examine which items contributed most information and to determine overall test reliability at specific levels of knowledge (transplant or CKD, as appropriate). We excluded items with very low discrimination values. In addition, we examined each item's fit to the model using the Orlando-Thissen χ^2 statistic.[4] Items with significant χ^2 values at $p<0.001$ were considered for exclusion due to poor fit. Results of these analyses are given below.

Detailed Differential Item Functioning Methods

Differential item functioning tests were conducted across race/ethnicity groups (non-Hispanic Black vs. other race/ethnicity), CKD stage (stage 3 vs. stages 4 and 5), primary language spoken (English vs. Spanish), and gender (woman vs. man). Differential item functioning indicates that membership in a specific patient group is associated with a systematically different item response probability. We examined differential item functioning using a hybrid logistic ordinal regression (LOR) item response theory approach. We used a McFadden pseudo-R2 change criterion of ≥ 0.02 to flag items for differential item functioning, and we utilized the lordif R package, version 0.3-3, for conducting the LOR DIF analyses.[5]

Results of Tests of IRT Assumptions

Unidimensionality. To determine whether the starting set of items was unidimensional, we fit a series of factor analysis models. First, we considered two formulations of the exploratory factor analysis model: the Schmid-Leiman (SL) transformation bifactor EFA model and the Jennrich-Bentler (JB) bifactor EFA model, each with their own, unique advantages. While both models include loadings on a unidimensional, general factor representing the overall construct, as well as multiple local factors representing potential sub-constructs or domains. The SL model is strictly an orthogonal model and has the advantage of generating more easily calculable statistics to determine the strength of the general factor, and therefore unidimensionality. These include McDonald's omega hierarchical reliability coefficient ($Omega_h$), which is the ratio of variance accounted for by the general factor to the total test variance. $Omega_h$ values of $\geq$0.70 indicate essential unidimensionality. The next is the explained common variance (ECV), which is defined as the ratio of the variance accounted for by the general factor to the total common variance. ECV values of $\geq$0.60 indicate essential unidimensionality. One drawback of the SL bifactor EFA model regards the lack of interpretability of any apparent local factors. On the other hand, the JB bifactor EFA model with oblique (correlated) rotation generates more interpretable local factors, but the calculation of omega hierarchical and ECV or more complex. Therefore, these models can work in combination to maximize each of their strengths. For each of the models, the items were entered after coding for correct vs. incorrect, and therefore dichotomous. Therefore, we used a tetrachoric correlation matrix for these analyses.

In the Schmid-Leiman (SL) transformation bifactor EFA model, most items had their highest loading on the general factor, with the exception of several items covering various aspects of the CKD knowledge CKD3, CKD4, CKD8, CKD9, CKD7, CKD6, CKD10, CKD5, CKD11), which loaded higher on a local factor. In addition, likely due to their similar structure, the two items asking about the likelihood of recipient and donor death during surgery (KTX25 and KTX20) loaded strongly on a local factor. With an $Omega_h$ of 0.75 and ECV of 0.60, the full item set met standards for unidimensionality. Table S2 shows this model's results. For the SL bifactor EFA, only factor loadings > 0.20 are shown.

However, because the local factors of an SL bifactor may be difficult to interpret, we sought to confirm them in a JB bifactor EFA. Indeed, similar pattern of strong loadings on local factors for some items relating to CKD knowledge (CKD3, CKD4, CKD8, CKD9, CKD7, CKD6, CKD10, CKD5, CKD11) emerged in this model as well, creating a parallel with the SL bifactor EFA. Similarly, the JB model showed strong loadings for KTX25 and KTX20 on a local factor, mirroring the results of the SL bifactor EFA. Table S3 shows this model's results.

Table S2. Schmid-Leiman Transformed Bifactor EFA Factor Loadings

Item	General Factor	Local Factor 1	Local Factor 2	Local Factor 3
CKD1	0.21			
CKD2	0.23			
CKD3	0.29		0.32	
CKD4	0.26		0.29	
KTX7	0.25			
CKD8	0.36		0.37	
CKD9	0.30		0.55	
CKD7	0.35		0.48	
CKD6	0.32		0.59	
CKD10	0.29		0.63	
CKD5	0.33		0.56	
CKD11	0.32		0.54	
KTX4	0.37			
KTX1	0.35			
KTX2	0.45			
KTX10	0.51			
KTX13	0.58			
KTX3	0.48			
KTX11	0.40			
KTX17	0.34			
KTX14	0.40			
KTX16	0.51			
KTX9	0.43			
KTX12	0.61			
KTX5	0.43			
KTX6	0.57			
KTX8	0.42			
KTX15	0.49			
KTX25	0.32			0.22
KTX19	0.21			
KTX24	0.22			
KTX18	0.40			
KTX22	0.51			
KTX23	0.22			
KTX26	0.22			0.67
KTX20	0.39			0.77
KTX21	0.34			

Table S3. Jennrich-Bentler Bifactor EFA Factor Loadings

	General Factor	Local Factor 1	Local Factor 2	Local Factor 3
CKD1	0.22	0.16	0.02	0.05
CKD2	0.26	0.18	-0.03	-0.04
CKD3	0.32	0.31	-0.07	0.05
CKD4	0.29	0.30	-0.05	0.02
KTX7	0.32	0.12	0.05	-0.25
CKD8	0.43	0.33	-0.09	-0.09
CKD9	0.35	0.51	-0.01	-0.01
CKD7	0.42	0.45	0.03	-0.12
CKD6	0.38	0.55	-0.02	-0.01
CKD10	0.34	0.59	0.02	0.03
CKD5	0.37	0.52	0.03	0.05
CKD11	0.37	0.50	0.00	0.03
KTX4	0.45	-0.07	-0.05	-0.29
KTX1	0.34	0.01	0.07	0.10
KTX2	0.47	0.00	0.02	-0.04
KTX10	0.50	-0.09	-0.04	0.08
KTX13	0.61	-0.12	0.04	-0.08
KTX3	0.43	0.00	0.01	0.29
KTX11	0.44	-0.05	0.02	-0.12
KTX17	0.44	-0.04	0.00	-0.37
KTX14	0.36	0.08	-0.01	0.34
KTX16	0.47	0.05	0.01	0.30
KTX9	0.37	0.02	-0.02	0.39
KTX12	0.56	-0.08	0.07	0.31
KTX5	0.39	0.01	-0.03	0.28
KTX6	0.55	-0.08	0.07	0.07
KTX8	0.41	-0.09	0.20	-0.02
KTX15	0.44	-0.07	-0.02	0.32
KTX25	0.29	-0.01	0.26	0.12
KTX19	0.21	0.02	0.15	-0.01
KTX24	0.21	-0.02	0.10	0.00
KTX18	0.39	-0.02	0.07	0.06
KTX22	0.52	0.02	0.12	0.00
KTX23	0.21	0.02	0.06	0.05
KTX26	0.19	0.03	0.66	-0.01
KTX20	0.35	-0.01	0.78	0.00
KTX21	0.35	0.04	0.11	-0.02

After examining the bifactor EFA models, we further tested the dimensional structure with confirmatory factor analysis (CFA) models. Since the items were dichotomous, we used the diagonally weighted least squares (DWLS) estimator. We compared the fit of two CFA models. First, we fit a unidimensional CFA model wherein all items loaded on one factor (Model 1). In a second model, specified two correlated factors, one representing CKD knowledge and one representing transplant knowledge (Model 2). Due to their very high correlation (r=0.89), and because they emerged as a doublet in the bifactor EFA analyses, we elected to omit one of the two items on the chance of recipients or donors dying during surgery (KTX26 and KTX20) from both CFA models. We tested the fit of CFA models with multiple criteria, including commonly-used fit indices: Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), and the root mean square error of approximation (RMSEA). Standard criteria for evaluating these indexes are as follows: RMSEA <.08, TLI > .95, and CFI > .95.[6] However, there is a growing recognition that these criteria are likely to be restrictive and may reject models that fit well.[7] Therefore, while taking these fit indices into consideration, we will take CFI values of >0.90 as a lower bound for acceptable fit. Table S4 shows the factor loadings for CFA Models 1 and 2.

Table S4. CFA Models 1 and 2 Factor Loadings

	Model 1	Model 2	
	Overall Factor	Transplant Knowledge	CKD Knowledge
KTX4	0.44	0.48	-
KTX1	0.43	0.47	-
KTX2	0.55	0.59	-
KTX10	0.58	0.63	-
KTX13	0.65	0.71	-
KTX3	0.57	0.61	-
KTX11	0.49	0.53	-
KTX17	0.44	0.48	-
KTX14	0.64	0.66	-
KTX16	0.65	0.69	-
KTX9	0.55	0.58	-
KTX12	0.69	0.74	-
KTX5	0.60	0.63	-
KTX6	0.65	0.71	-
KTX8	0.54	0.58	-
KTX15	0.62	0.66	-
KTX25	0.46	0.50	-
KTX19	0.35	0.37	-
KTX24	0.31	0.34	-
KTX18	0.52	0.56	-
KTX22	0.63	0.68	-
KTX23	0.49	0.51	-
KTX21	0.51	0.54	-
KTX20	0.50	0.55	-
KTX7	0.39	0.40	-
CKD1	0.34	-	0.39
CKD2	0.43	-	0.49
CKD3	0.51	-	0.59
CKD4	0.50	-	0.58
CKD8	0.62	-	0.70
CKD9	0.64	-	0.74
CKD7	0.80	-	0.88
CKD6	0.70	-	0.80
CKD10	0.69	-	0.78
CKD5	0.68	-	0.77
CKD11	0.66	-	0.75

Though its model chi-square was significant [2483.42 (df=594), p<0.001], Model 1's fit indexes revealed reasonable fit with RMSEA = 0.06, CFI = 0.92, and TLI = 0.91. The large majority loaded on an overall factor at >0.40. However, while Model 2's chi-square was also statistically significant [1335.91 (df=593), p<0.001], its fit was superior to that of Model 1 on the fit indexes with RMSEA = 0.04, CFI = 0.97, and TLI = 0.97. The two factors in Model two were highly correlated at 0.60. In addition to the superior of fit of Model 2, loadings for the CKD items were somewhat higher on an independent, CKD factor in Model 2 than they were on an overall knowledge factor in Model 1. Therefore, we concluded that a better representation of the data would be two separate scales for CKD and Transplant knowledge. We proceeded with IRT modeling of these sets of items independently.

Monotonicity. Figures S1a and S1b below show item characteristic curves for each item for the 2PL IRT models examining transplant knowledge and CKD knowledge items (separate models).

Figure S1a. Item Characteristic Curves for Transplant 2PL Model

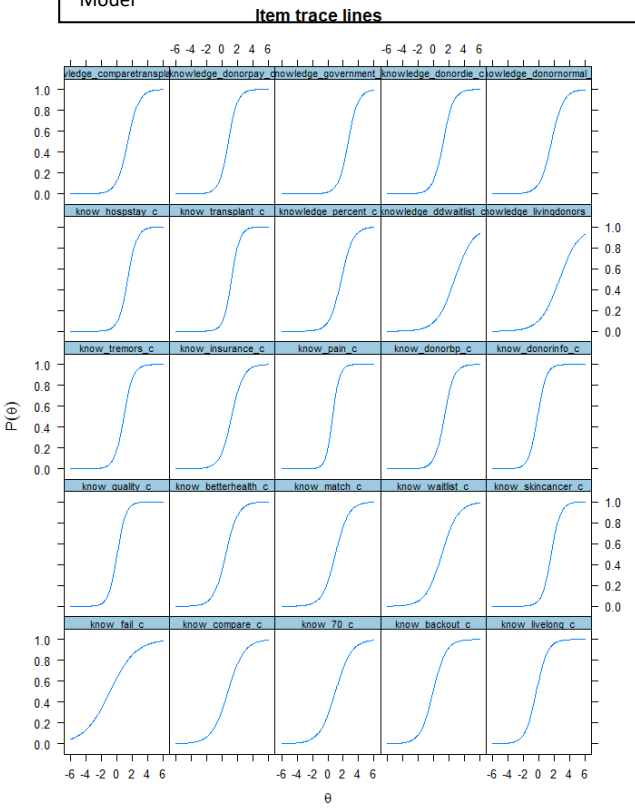
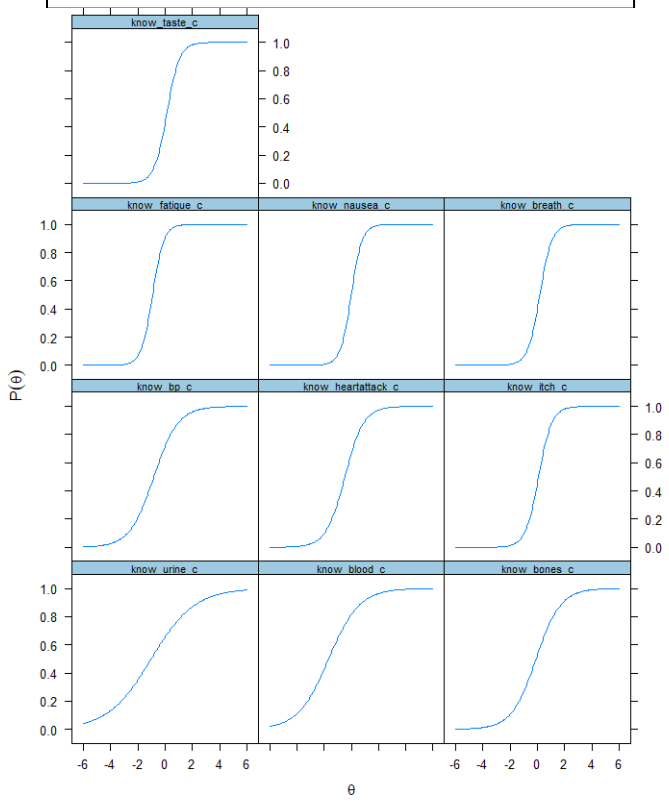


Figure S1b. Item Characteristic Curves for CKD 2PL Model



Exclusions from the Initial Round of IRT Analyses

Transplant Knowledge Items

Six items were omitted for low discrimination values: KTX7 (0.60), KTX4 (0.95), KTX17 (0.95), KTX19 (0.85), and KTX24 (0.79). Only one additional item, KTX2, had a significant Orlando-Thissen χ^2 statistic (χ^2 = 36.08 (df=16), p=0.003) and was omitted on this basis. Finally, 3 items (KTX3, KTX16, KTX20) exhibited local dependence with several other items and were omitted. Local dependence was examined with the Jackknife Slope Index, which quantifies the extent of change in slope parameters when sequentially removing each item.

CKD Knowledge Items

Though a few items had low discrimination values [CKD1 (0.61) and CKD2 (0.84)] and significant Orlando-Thissen χ^2 statistics [CKD4: (χ^2 = 17.77 (df=8), p=0.02); CKD7: (χ^2 = 15.45 (df=6), p=0.02); CKD10: (χ^2 = 16.93 (df=7), p=0.02)], these were kept for the clinical importance of their content after review. However, CKD6 exhibited strong local dependence with CKD5, so we elected to omit CKD6.

Results of Second Round of IRT Analyses

Transplant Knowledge Model. After the initial round of item response theory analyses, we refit a separate two-parameter logistic model for the transplant knowledge items (16 remaining items; Table S5a.) The *a* (discrimination) parameter exceeded 1 in all but 1 item (KTX1, "Patients older than 70 years can receive transplants."), indicating that most items adequately distinguished between patients with higher and lower transplant knowledge. Regarding *b* estimates, the easiest item to answer correctly was KTX10 ("In general, patients can live longer with a kidney transplant than if they stayed on dialysis.") and the most difficult item was KTX23 ("After a transplant, how long does the US Government pay for most of the costs of transplant medications?"). None of the items met the stricter criterion of $p<0.001$ for item misfit on the Orlando-Thissen χ^2. Across the range of theta for these 16 items, reliability was highest between theta values of 0 and 1 (0 indicates population mean and 1 is a standard deviation better knowledge than the population mean), with a marginal reliability of 0.84. Reliability was especially low beyond -2 and 3. Based on these results, we elected not to exclude any of the transplant knowledge items. As legislation was passed to provide lifetime Medicare coverage for immunosuppressants in December 2020, the correct answer to KTX23 ("After a transplant, how long does the US Government pay for most of the costs of transplant medications?"), will be changed from "three years" to "for the patient's entire life."

Table S5a. Parameters from Two-Parameter Logistic (2PL) Item Response Theory (IRT) Model for Transplant Knowledge Items

Item	Discrimination (*a*)[1]	Difficulty (*b*)[2]	Factor Loading	SX2 (df), p-value[3]
KTX1	0.94	1.02	0.48	6.60 (10), 0.75
KTX10	1.34	-0.30	0.62	11.38 (9), 0.25
KTX13	1.58	0.00	0.68	7.81 (9), 0.55
KTX11	1.07	1.07	0.53	13.61 (10) 0.19
KTX14	1.64	1.69	0.70	11.12 (9), 0.27
KTX9	1.46	1.14	0.65	9.37 (10), 0.50
KTX12	2.21	0.64	0.79	11.18 (9), 0.26
KTX5	1.57	1.46	0.66	8.76 (10) 0.56
KTX6	1.66	-0.16	0.70	10.48 (8), 0.23
KTX8	1.49	1.47	0.66	19.82 (10), 0.03
KTX15	1.73	1.24	0.71	13.23 (10), 0.21
KTX25	1.23	1.85	0.59	9.63 (11), 0.56
KTX18	1.40	1.40	0.63	5.00 (10), 0.89
KTX22	1.62	0.78	0.69	7.29 (10), 0.70
KTX23	1.43	2.81	0.64	11.74 (9), 0.23
KTX21	1.24	1.77	0.59	11.15 (11), 0.43

[1] The discrimination parameter represents how well the item differentiates between patients with higher and lower levels of knowledge. This parameter is on a metric with population mean of 0 and standard deviation of 1 with higher values indicating better discrimination.

[2] The difficulty parameter captures the how difficult it is to give a correct answer to the question. This parameter is on a metric with population mean of 0 and standard deviation of 1 with higher values indicating more difficult items.

[3] The p-value for the Orlando-Thissen $\chi 2$ statistic (SX2) indicates whether the item is a good fit (or not) to the item response theory model. P-values <0.001 indicate poor fit.

CKD Knowledge Model. After making item exclusions from the initial round of item response theory analyses, we refit separate two-parameter logistic item response theory models for the CKD knowledge items (10 remaining items). A table of these results is given below. (Table S5b.) With the exception of CKD1 ("Does the kidney make urine?"), all items had reasonable discrimination values, with several exceeding 2, indicating that most items distinguished between patients with higher and lower CKD knowledge. In comparison to the transplant knowledge items, the CKD knowledge items tended to be easier for participants to answer correctly, with more *b* parameters below 0. The easiest item was CKD2 ("Does the kidney make clean blood?") and the hardest item was CKD5 ("Shortness of breath is a symptom of chronic kidney disease."). None of the items showed misfit using the stringent criterion. We omitted two CKD knowledge items, CKD1 ("Does the kidney make urine?") and CKD6 ("Shortness of breath is a symptom of chronic kidney disease"). There was strong local dependence between CKD6 and CKD5 ("Shortness of breath is a symptom of chronic kidney disease")]. We elected to omit CKD6 since it is a less common symptom of CKD. For these 9 items, reliability was high between one standard deviation below and above the mean (-1 to 1), with a marginal reliability of 0.91 in this range. However, it declined outside these ranges.

Table S5b. Parameters from Two-Parameter Logistic (2PL) Item Response Theory (IRT) Model for CKD Knowledge Items

Item	Discrimination (*a*)[1]	Difficulty (*b*)[2]	Factor Loading	SX2 (df), p-value[3]
CKD1	0.63	-1.01	0.35	10.24 (7), 0.18
CKD2	0.89	-1.65	0.46	8.00 (7), 0.33
CKD3	1.12	-0.10	0.55	8.19 (7), 0.32
CKD4	1.13	-0.82	0.55	14.59 (7), 0.04
CKD8	1.46	-0.47	0.65	3.58 (7), 0.83
CKD9	2.08	0.09	0.77	9.25 (6), 0.16
CKD7	2.45	-0.92	0.82	15.33 (6), 0.02
CKD10	2.52	0.00	0.83	11.42 (6), 0.08
CKD5	2.29	0.18	0.80	4.56 (6), 0.60
CKD11	2.18	0.14	0.79	7.82 (6), 0.25

[1] The discrimination parameter represents how well the item differentiates between patients with higher and lower levels of knowledge. This parameter is on a metric with population mean of 0 and standard deviation of 1 with higher values indicating better discrimination.

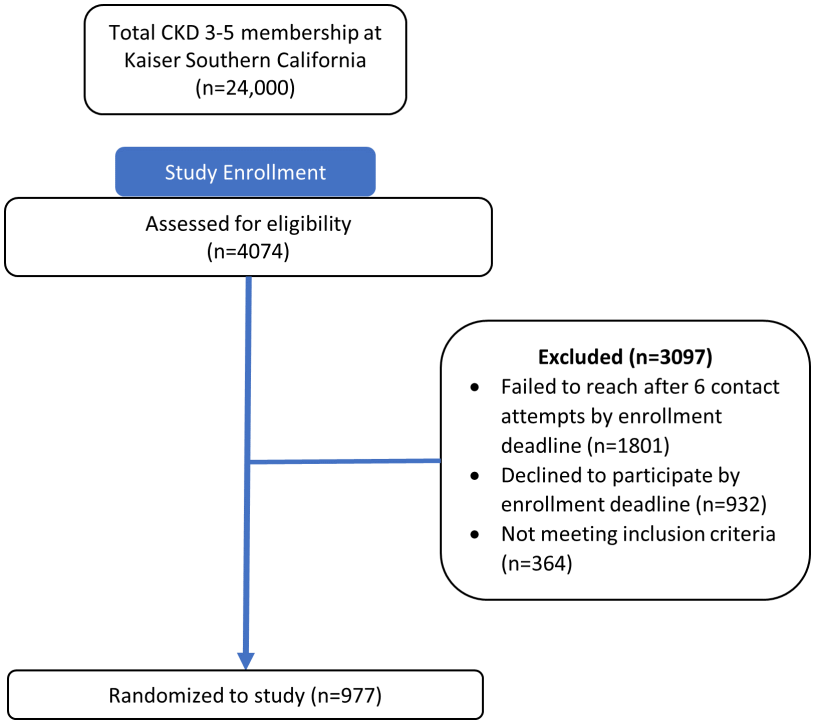
[2] The difficulty parameter captures the how difficult it is to give a correct answer to the question. This parameter is on a metric with population mean of 0 and standard deviation of 1 with higher values indicating more difficult items.

[3] The p-value for the Orlando-Thissen $\chi 2$ statistic (SX2) indicates whether the item is a good fit (or not) to the item response theory model. P-values <0.001 indicate poor fit.

Detailed Results of Differential Item Functioning Analyses

No items flagged for differential item functioning across groups of race/ethnicity, CKD stage, and gender based on McFadden pseudo-R2 change criterion of ≥ 0.02. One item, KTX9 ("All living donors transportation costs to and from the transplant center are covered by Medicare and/or private insurance") exceeded the McFadden pseudo-R2 change criterion by a very small amount (<0.02). However, since no confusion was expressed by Spanish-speaking patients about this item during cognitive interviews, we elected not to exclude it. Therefore, we did not exclude any items due to differential item functioning.

Supplemental Figure 1. Patient Recruitment and Inclusion Flowchart



References

1. Waterman AD, Stanley SL, Covelli T, Hazel E, Hong BA, Brennan DC. Living Donation Decision Making: Recipients' Concerns and Educational Needs. 2006;16(1):17-23.
2. Waterman AD, Barrett AC, Stanley SL. Optimal transplant education for recipients to increase pursuit of living donation. *Progress in transplantation.* 2008;18(1):55-62.
3. Chalmers RP. mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software.* 2012;48(6):1-29.
4. Orlando M, Thissen D. Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Appl Psych Meas.* 2000;24(1):50-64.
5. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *J Stat Softw.* 2011;39(8):1-30.
6. Hu LT, Bentler PM. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Struct Equ Modeling.* 1999;6(1):1-55.
7. Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. *Qual Life Res.* 2009;18(4):447-460.