The eGVHD App has the potential to improve the accuracy of graft-*versus*-host disease assessment: a multicenter randomized controlled trial

Helene M. Schoemans,[1,2] Kathy Goris,[1] Raf Van Durm,[3] Steffen Fieuws,[4] Sabina De Geest,[2,5] Steven Z. Pavletic,[6] Annie Im,[7] Daniel Wolff,[8] Stephanie J. Lee,[9] Hildegard Greinix,[10] Rafael F. Duarte,[11] Xavier Poiré,[12] Dominik Selleslag,[13] Philippe Lewalle,[14] Tessa Kerre,[15] Carlos Graux,[16] Frédéric Baron,[17] Johan A. Maertens[1] and Fabienne Dobbels;[2] on behalf of the EBMT Transplantation Complications Working party

[1]Department of Hematology, University Hospitals Leuven and KU Leuven, Belgium; [2]Academic Centre for Nursing and Midwifery, KU Leuven, Belgium; [3]IT Department, University Hospitals Leuven, KU Leuven, Belgium; [4]L-BioStat, KU Leuven – University of Leuven & Universiteit Hasselt, Leuven, Belgium; [5]Institute of Nursing Science, Department Public Health, University of Basel, Switzerland; [6]Experimental Transplantation and Immunology Branch, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD, USA; [7]University of Pittsburgh Medical Center, Pittsburgh, PA, USA; [8]Department of Hematology and Clinical Oncology, University of Regensburg, Regensburg, Germany; [9]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [10]Division of Hematology, Medical University of Graz, Austria; [11]ICO/Hospital Duran I Reynals, Hospitalet De Llobregat, Spain; [12]Cliniques Universitaires Saint-Luc, Brussels, Belgium; [13]Department of Hematology, AZ Sint-Jan Brugge, Brugge, Belgium; [14]Institut Jules Bordet - Université Libre de Bruxelles, Belgium; [15]Hematology and Stem Cell Transplantation, Ghent University Hospital, Belgium; [16]Université Catholique de Louvain, CHU UCL Namur (Godinne site), Yvoir, Belgium and [17]Hematology, University of Liège, GIGA-I3, Belgium

 doi:10.3324/haematol.2018.190777

Received: March 6, 2018.

Accepted: June 13, 2018.

Pre-published: August 9, 2018.

Correspondence: helene.schoemans@uzleuven.be

Supplemental methods

Sample and setting – Workshop description

The workshop lasted 90 to 120 minutes. It consisted of an introduction to the eGVHD App project, a short tutorial to train the "APP" group in the basic functionalities of the eGVHD App and the completion of a demographics, practice patterns and technology acceptance survey ("Survey 1") by all participants. Both groups then received the GvHD test package and recorded their answers individually. This was followed by a usability survey ("Survey 2") restricted to the "APP" group. The workshop concluded with a discussion of the correct answers of the test package and a summary of the most current recommendations for GvHD assessment.

Data collection points, randomization procedure and blinding

All data collection was performed during the workshop using pen and paper. Allocation to the intervention arm ("APP" group) was random and stratified. More specifically, randomization was done at the arrival of study participants based on pre-formatted randomization sheets (www.randomization.com) and order of arrival. We used randomly permuted blocks, with block sizes of 2, to compensate for the low number of participants per center. Stratification was based on professional background: (1) senior physicians (board certified hematologists), (2) junior physicians (medical doctors training in internal medicine or hematology), (3) data managers or research nurses specialized in HCT data entry or (4) other (e.g. medical students or nurses with no specific GvHD evaluation expertise). Blinding was not feasible due to the nature of the intervention.

Outcome measures – Planned sub-analyses

We planned the following sub-analyses: (1) to compare the difference in diagnosis accuracy between both groups, (2) to test for the App-effect on the accuracy of the severity scoring conditional on a GvHD diagnosis being acute (aGvHD) and chronic (cGvHD), (3) to verify whether the effect of the App depended on the type of GvHD, the severity of GvHD, professional background or center, (4) to compare the inter-rater reliability and (5) to compare the time needed to complete the full test package between both groups.

Variables and measurements

Gold Standard

Prior to this study, four GvHD experts (SZP, DW, AI and SJL) determined the correct diagnosis and severity score of each vignette, based on the MAGIC criteria[1] for acute GvHD and the NIH 2014 guidelines[2] for chronic GvHD by evaluating the ten clinical vignettes independently and returning their GvHD assessment separately to the principal investigator. The 'gold standard' for diagnosis and severity scoring corresponded to the answer given by at least three of the four experts. When an expert disagreed with the consensus of the other three experts, this expert was contacted separately to confirm that he/she agreed with the 'gold standard' answer given by the rest of the group.

Experts were healthcare professionals active in the field of allogeneic HCT, co-authors of at least one publication in the field of clinical GvHD and active members of an international GvHD consortium or working group.

Post-test user satisfaction and experience

Briefly, the TAM consists of six statements, referring to the extent to which the user believes the technology will improve his work performance. Statements are rated on a 7-point Likert-like scale (1= 'extremely unlikely' to 7= 'extremely likely'). A median score is calculated for each item separately, with higher scores reflecting higher perceived usefulness. The PSSUQ is a 19-item questionnaire using 7-point Likert-like scales (1= 'strongly agree' to 7= 'strongly disagree'), with three subscales reflecting system usefulness (items 1-8), information quality (items 9-15) and interface quality (items 16-18), respectively. PSSUQ scores are presented as median total and subscale scores, with lower scores reflecting higher user satisfaction.

Statistical analysis

Results were analyzed with IBM SPSS statistics version 24 and R version 3.3.3 according to the 'intention to treat' principle. Missing results were reported as such. Descriptive results were reported using a measure of central tendency and a measure of dispersion, as appropriate. The probability of a correct answer was compared between both groups using a mixed effects logistic regression model, for diagnosis and severity score separately. The model contained *fixed effects* of group ("App" versus "No App") and professional background (the stratification variable in the randomization) and *random effects* of center and workshop participant. These random effects were included to handle the correlation between the workshop

participants belonging to the same center, and between the ten answers given by the same workshop participant, respectively. Odds ratio's and 95% confidence intervals (CI) for the effect of group were reported. To verify whether the effect of the App depended on the type of the GvHD, the severity of the GvHD, professional background or center, interaction terms were added in separate models. *Inter-observer agreement* of the severity was evaluated by using the Brennan-Prediger's kappa coefficient (K_{BP}) which ranges between zero (no agreement) and one (perfect agreement). This coefficient evaluates the raters' agreement for nominal scales with more than two categories and takes into account the fact that agreement could have occurred by chance. This version of the kappa is reported instead of the classical Fleiss-Cohen kappa, since the latter is not appropriate for comparisons of conditions having a difference in distribution[3]. Kappa's are compared between both groups using an approach presented by Gwet and colleagues[4]. The *time needed to score* the vignettes was compared between both groups using a linear mixed model, with the same fixed and random effects as in the aforementioned logistic regression model.

Supplemental Tables

Supplementary Table 1:

Characteristics of Centers Performing Allogeneic HCT in Belgium and Participating in the Study

Center	Academic Center	total HCT per year	1st alloHCT per year	% activity in Belgium	total number of participants	Senior MD	Junior MD	Data managers	Other
1	yes	130	76	82%	20	8	8	4	0
2	yes	95	57		13	5	4	3	1
3	no	87	33		8	3	2	1	2
4	yes	74	34		7	4	1	1	1
5	yes	79	49		13	6	3	4	0
6	yes	57	20		8	7	0	1	0
7	yes	93	43		8	4	3	1	0
8	no	39	17	18%	Declined	NA	NA	NA	NA
9	yes	47	26		Declined	NA	NA	NA	NA
10	yes	31	12		Declined	NA	NA	NA	NA
11	no	35	15		Declined	NA	NA	NA	NA

LEGEND

HCT: hematopoietic stem cell transplantation; alloHCT: allogeneic hematopoietic stem cell transplantation; MD: medical doctor

Supplementary Table 2: List of Clinical Vignettes

Vignette	Decription of the Clinical Vignette	Diagnosis	Severity Scoring
1	1. A female adult patient receives an allogeneic stem cell transplantation for a myelodysplasia. Her post transplantation period is uneventful, but 9 months after transplantation she develops: • a red inflammatory rash on both arms, one month after discontinuation of immunosuppression. • There are no other abnormal signs or symptoms and her pulmonary function lab results are normal. • A biopsy of the skin of the forearm is suggestive for GVHD (likely GVHD - apoptosis in epidermal basal layer).	Late acute GVHD	Grade I
2	A female adult patient receives an allogeneic stem cell transplantation for a chronic myeloid leukemia. Her pre-transplantation evaluation is unremarkable. Around day 90, she develops: • dyspnea when walking on flat ground. • A pulmonary evaluation reveals a newly decreased FEV1* of 65%, with a FEV1/VC ratio** of 0.65 and a RV*** of 110%. • Air trapping is present on high resolution CT scan of the lungs. • Infections of the respiratory tract are excluded by a normal bronchial aspirate evaluation and her cardiac function is normal. • Her clinical exam and laboratory results are perfectly normal except for xerostomia (dry mouth), without impact on her oral intake.	Classic Chronic GVHD	Moderate

3	Four months after receiving an allogeneic stem cell transplantation, a female adult patient presents with: • anorexia, daily vomiting and an unintentional weight loss of about 15% of her pre-transplantation weight. • There are no other abnormal signs or symptoms and her lab results and pulmonary function tests are normal. • A stomach biopsy confirms GVHD (likely GVHD - gastric pit apoptosis).	Late acute GVHD	Grade II
4	Four weeks after receiving an allogeneic stem cell transplantation, a male adult patient presents with: • an itchy erythematous rash involving the head and neck, and anorexia with major diarrhea (10x/day, about 2000ml/day) but no abdominal pain. • A colonoscopy confirms GVHD by biopsy (likely GVHD - crypt apoptosis in the intestines) and excludes a concomitant infection or drug toxicity. • His lab results are normal except for a low albumin and slightly elevated creatinine. • His pulmonary function tests are normal.	Classic acute GVHD	Grade III
5	A female adult patient, 6 months after her allogeneic stem cell transplantation, develops: • two patches of morphea-like lesions (patches of leather-like, shiny skin) on the lower back (diameter 5cm) • with an elevation of liver enzymes (ALT, AST, AP and GGT a little more than 3x the upper normal limit), without other potential confounding cause. • Her pulmonary function tests are normal. • She reports dyspareunia (painful intercourse) and a gynecological exam reveals vaginal adhesions and scarring.	Overlap Chronic GVHD	SEVERE

6	Two months after receiving an allogeneic stem cell transplantation, a male adult patient develops: • relatively frequent diarrhea episodes (4 times/ a day) accompanied with very severely painful abdominal cramps. • A colonoscopy confirms GVHD by biopsy (likely GVHD – apoptosis in enterocytes and destruction of crypt architecture) and excludes a concomitant infection / drug toxicity. • His body weight is unchanged. • Liver enzymes are slightly elevated (ALT, AST, AP and GGT slightly more than twice the upper normal limit) and bilirubin is 3.5 mg/dL, without argument for infection, drug toxicity or veno-occlusive disease. • Except for fatigue, there are no other abnormal signs or symptoms and the rest of his lab results are normal. • His pulmonary function tests are normal.	Classic acute GVHD	Grade IV

7	A male adult patient receives an allogeneic stem cell transplantation. Ten months later, he feels fine but he reports: • frequent muscle cramps and has noticed that some movements have become more difficult : an increasing tightness in his lower back, arms and legs is making it more difficult to his daily jog and pick up items on the ground. • On clinical exam, the extension of the arms and the flexion of the wrist are somewhat decreased and the ankles moderately swollen without inflammatory features. • No other clinical abnormalities are found than new lichen sclerosus-like changes (white patches of firm thickened/crinkled skin with a tendency to scar) that have appeared on the penis. • An electromyography is normal. His laboratory exams are normal, including muscle enzymes. • His pulmonary function tests are normal.	Classic Chronic GVHD	Moderate
8	Three months after her transplantation, a female adult allogeneic transplantation recipient presents with: • two new painful ulcerations in the mouth. Oral exam reveals lichen planus like changes and 1.5cm wide ulcerations. Microbial examination for candida and herpes are negative. • Weight is stable but the patient no longer tolerates sparkling drinks. Oral intake is preserved. • The rest of her clinical exam, pulmonary function tests and lab results are unremarkable.	Classic Chronic GVHD	MILD
9	A male adult patient, five months after receiving an allogeneic stem cell transplantation, develops: • a new maculopapular inflammatory red rash on the hands and feet.	Overlap Chronic GVHD	MILD

	• He also notices that his nails have become brittle and his eyes are more sensitive than before. • The ophthalmologist confirms signs of keratoconjunctivitis sicca with a slit lamp examination. Shirmer's test shows a 3mm tear production after 5 minutes. • His ocular problems are totally relieved by using artificial teardrops twice a day. • Further exams reveal normal pulmonary function tests, an unremarkable clinical exam (except for the rash and dystrophic nails). • He has normal laboratory results except for slightly elevated alkaline phosphatase (AP) which are a little over twice the upper normal limit (2x ULN).		
10	Twelve months after her allogeneic stem cell transplantation, a female adult patient develops: • difficulties with swallowing due to non-painful xerostomia (dry mouth) and the impression that food remains stuck when she swallows. • Her weight remains stable but she needs to chew abnormally long and drink along almost all of her solid food intakes. • A gastroscopy confirms the presence of a new stenosis of the upper esophagus, which is successfully dilated but no biopsies are taken. • Her other clinical, laboratory and pulmonary function test evaluations are normal, except for some superficial sclerosis bilaterally in the lower arms and legs.	Classic Chronic GVHD	SEVERE

LEGEND

* FEV1 = forced expiratory volume in one second

**FEV1/VC =Tiffeneau index or forced expiratory volume in one second divided by vital capacity

***RV= residual volume

Supplementary Table 3: Full Binary (Correct vs Incorrect) Results of Participants for Individual GvHD vignettes

LEGEND

* missing results were considered as being incorrect; v.s.: versus; Δ: difference between

Supplementary Table 3: **Full Binary (correct vs. incorrect) Results of Participants for Individual GvHD Vignettes**

Vignette number	Gold Standard Diagnosis	Participant Diagnosis*	All participants n=77	"APP" group n=37	"NO APP" group n=40	Δ "App" vs. "No App" (%) Correct Diagnosis	Gold Standard Severity Scoring	Participant Severity Scoring	All participants n=77	"APP" group n=37	"NO APP" group n=40	Δ "App" vs. "No App" (%) Correct Severity Scoring
1	Late **acute** GVHD	correct	69 (89.6%)	33 (89.2%)	36 (90%)	- 0.8 %	Grade I	correct	62 (80.5%)	33 (89.2%)	29 (72.5%)	+ 16.7 %
		incorrect	8 (10.4%)	4 (10.8%)	4 (10%)			incorrect	15 (19.5%)	4 (10.8%)	11 (27.5%)	
3	Late **acute** GVHD	correct	60 (77.9%)	37 (100%)	23 (57.5%)	+ 42.5%	Grade II	correct	48 (62.3%)	37 (100%)	11 (27.5%)	+ 72.5%
		incorrect	17 (22.1%)	0 (0%)	17 (42.5%)			incorrect	29 (37.7%)	0 (0%)	29 (72.5%)	
4	Classic **acute** GVHD	correct	72 (93.5%)	34 (91.9%)	38 (95%)	- 3.1%	Grade III	correct	62 (80.5%)	35 (94.6%)	27 (67.5%)	+ 27.1%
		incorrect	5 (6.5%)	3 (8.1%)	2 (5%)			incorrect	15 (19.5%)	2 (5.4%)	13 (32.5%)	
6	Classic **acute** GVHD	correct	75 (97.4%)	37 (100%)	38 (95%)	+ 5%	Grade IV	correct	40 (51.9%)	33 (89.2%)	7 (17.5%)	+ 71.7%
		incorrect	2 (2.6%)	0 (0%)	2 (5%)			incorrect	37 (48.1%)	4 (10.8%)	33 (82.5%)	
8	Classic **Chronic** GVHD	correct	64 (83.1%)	35 (94.6%)	29 (72.5%)	+ 22.1%	Mild	correct	39 (50.6%)	19 (51.3%)	20 (50%)	+ 1.3%
		incorrect	13 (16.9%)	2 (5.4%)	11 (27.5%)			incorrect	38 (49.4%)	18 (48.6%)	20 (50%)	
9	Overlap **Chronic** GVHD	correct	68 (88.3%)	36 (97.3%)	32 (80%)	+ 17.3%	Mild	correct	42 (54.5%)	30 (81.1%)	12 (30%)	+ 51.1%
		incorrect	9 (11.7%)	1(2.7%)	8 (20%)			incorrect	35 (45.5%)	7 (18.9%)	28 (70%)	
2	Classic **Chronic** GVHD	correct	66 (85.7%)	35 (94.6%)	31 (77.5%)	+ 17.1%	Moderate	correct	48 (62.3%)	32 (86.5%)	16 (40%)	+ 46.5%
		incorrect	11 (14.3%)	2 (5.4%)	9 (22.5%)			incorrect	29 (37.7%)	5 (13.5%)	24 (60%)	
7	Classic **Chronic** GVHD	correct	72 (93.5%)	36 (97.3%)	36 (90%)	+ 7.3%	Moderate	correct	53 (68.8%)	29 (78.4%)	24 (60%)	+ 18.4%
		incorrect	5 (6.5%)	1 (2.7%)	4 (10%)			incorrect	24 (31.2%)	8 (21.6%)	16 (40%)	
5	Overlap **Chronic** GVHD	correct	72 (93.5%)	37 (100%)	35 (87.5%)	+ 12.5%	Severe	correct	41 (53.2%)	29 (78.4%)	12 (30%)	+ 48.4%
		incorrect	5 (6.5%)	0 (0%)	5(12.5%)			incorrect	36 (46.7%)	8 (21.6%)	28 (70%)	
10	Classic **Chronic** GVHD	correct	74 (96.1%)	37 (100%)	37 (92.5%)	+ 7.5%	Severe	correct	40 (51.9%)	25 (67.6%)	15 (37.5%)	+ 30.1%
		incorrect	3 (3.9%)	0 (0%)	3 (7.5%)			incorrect	37 (48.1%)	12 (32.4%)	25 (62.5%)	

LEGEND

* missing results were considered as being incorrect

v.s.: versus; Δ: difference between

Supplementary Table 4: Post-test User experience and Usability Data ("APP" group only)

LEGEND

IQR: inter quartile range; TAM: Technology Assessment Model; PSSUQ: Post-Study System Usability Questionnaire

Supplementary Table 4: Post-test User Experience and Usability Data ("APP" group only)

Perceived Usefulness – TAM median score (**7= extremely likely**; 1= extremely unlikely)		
Using the "EBMT GVHD app" would…	n	
Enable me to accomplish tasks more quickly.	37	5 (IQR 2; range: 2-7)
Improve my job performance.	37	6 (IQR 1; range: 5-7)
Increase my productivity.	37	5 (IQR 1; range: 3-7)
Enhance my effectiveness on the job.	37	6 (IQR 1; range: 4-7)
Make it easier to do my job.	37	6 (IQR 1; range: 3-7)
I would find the "EBMT GVHD app" useful in my job.	37	6 (IQR 1; range: 4-7)
System Usability – PSSUQ median score (**1= Strongly agree** ; 7= Strongly disagree)	n	
1. Overall, I am satisfied with how easy it is to use this system.	36	2 (IQR 1; range 1-3)
2. It was simple to use this system.	36	2 (IQR 1; range 1-5)
3. I could effectively complete the tasks and scenarios using this system.	36	2 (IQR 1; range 1-4)
4. I was able to complete the tasks and scenarios quickly using this system	36	3 (IQR 1; range 1-6)
5. I was able to efficiently complete the tasks and scenarios quickly using this system.	36	2 (IQR 1; range 1-6)
6. I felt comfortable using this system.	36	2 (IQR 2; range 1-5)
7. It was easy to learn to use this system.	36	1.5 (IQR 1; range 1-5)
8. I believe I could become productive quickly using this system.	36	2 (IQR 1; range 1-5)
System use subscale score		2 (IQR 1; range 1-5)
9. The system gave error messages that clearly told me how to fix problems.	26	2.5 (IQR2; range 1-6)
10. Whenever I made a mistake using the system, I could recover easily and quickly.	34	2 (IQR 1; range 1-5)
11. The information provided with this system was clear.	36	2 (IQR 0; range 1-3)
12. It was easy to find the information I needed.	29	2 (IQR 0; range 1-5)
13. The information provided for the system was easy to understand.	35	2 (IQR: 0; range 1-2)
14. The information was effective in helping me complete the tasks and scenarios.	35	2 (IQR: 0; range 1-3)
15. The organization of information on the system screens was clear.	36	2 (IQR: 1; range 1-6)
Information quality subscale score		2 (IQR 1; range 1-3)
16. The interface (= items you use to interact with the system e.g. screen, mouse, keyboard,…) of this system was pleasant.	35	2 (IQR: 1; range 1-6)
17. I liked using the interface of this system.	35	2 (IQR: 1; range 1-5)
18. This system has all the functions and capabilities I expect it to have.	36	2 (IQR: 1; range 1-3)
Interface quality subscale score		2 (IQR 1; range 1-3)
19. Overall, I am satisfied with this system.	36	2 (IQR: 1; range 1-3)
Overall PSSUQ (items 1-19)		2 (IQR 0,4; range 1-3)
Predicted use		
Reported level of **likelihood of using the app** in the future (Likert scale 1 (lowest)-10 (highest))	36	8 (IQR 3; range: 1-10)
Actual use		
Reported level of **comfort using the app** in English (Likert scale 1 (lowest)-10 (highest))	37	9 (IQR 2.5; range: 3-10)

LEGEND
IQR: inter quartile range; TAM: Technology Assessment Model; PSSUQ: Post-Study System Usability Questionnaire

Supplemental Figures

Supplementary Figure 1 – The superiority of the eGVHD App in GvHD assessment is similar regardless of **center effect** for diagnosis (A) and severity scoring (B)

Supplementary Figure 2 – The superiority of the eGVHD App in GvHD assessment is similar regardless of **professional background** for diagnosis (A) and severity scoring (B)

Supplementary Figure 3 – The superiority of the eGVHD App in GvHD assessment is similar regardless of the **age category** of the user for diagnosis (A) and severity scoring (B)

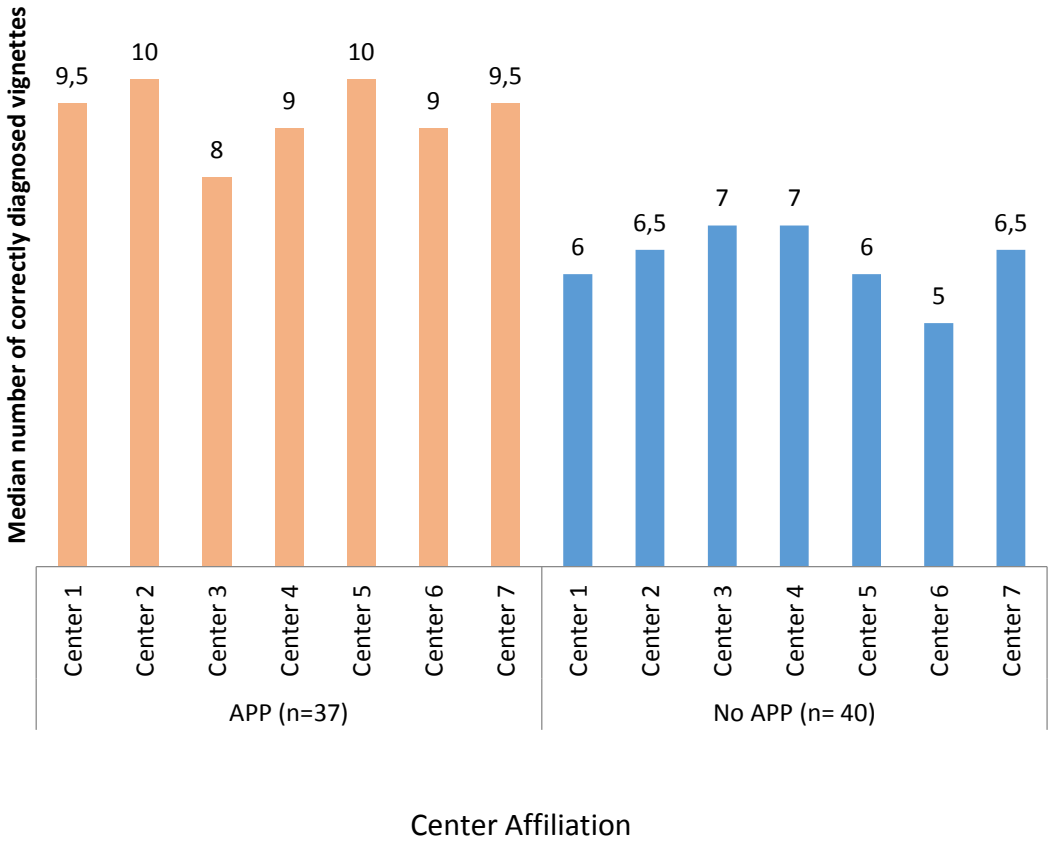
Supplementary Figure 4 – The superiority of the eGVHD App in GvHD assessment is similar regardless of the **user self-reported experience with GvHD assessment** for diagnosis (A) and severity scoring (B)

Legend supplementary Figure 4:
User experience with GvHD assessment was categorized according to the number of HCT patients the health care professionals reported to evaluate per week: "very low" experience (less than one HCT patient per week), "low" (1 to 6 weekly contacts with HCT patients), "moderate" experience (7 to 15 weekly contacts with HCT patients); "high" experience (more than 15 weekly contacts with HCT patients)

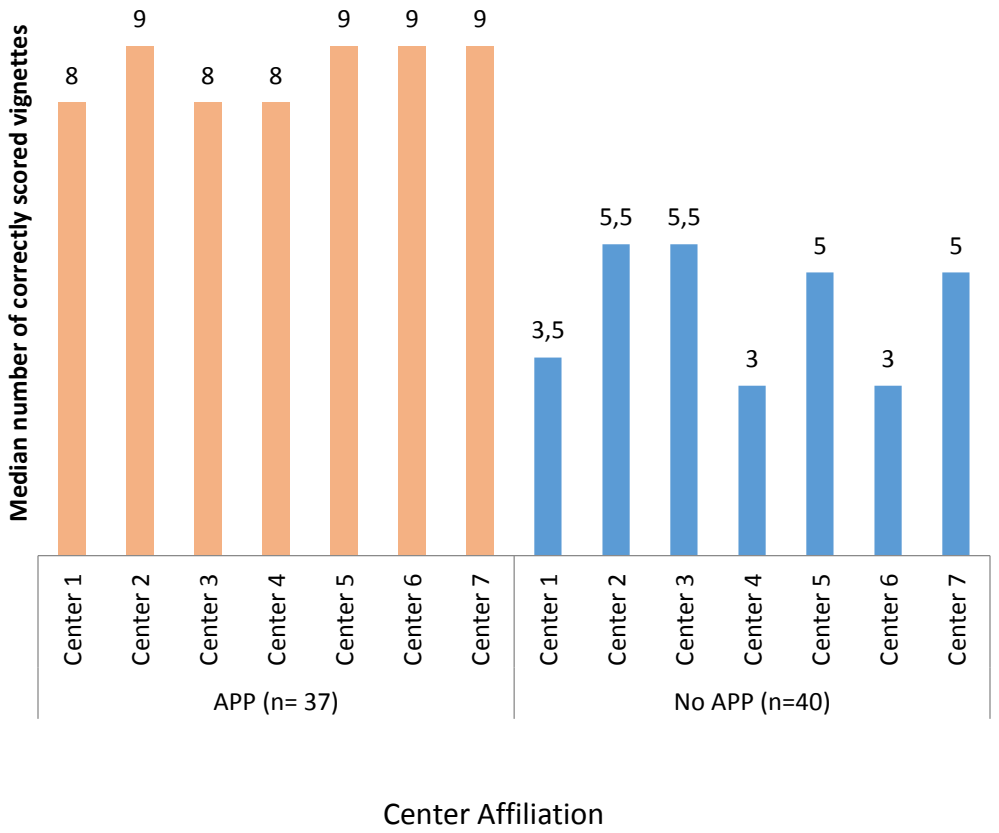
Supplementary Figure 5 – The superiority of the eGVHD App in GvHD assessment is similar regardless of the **user self-reported comfort with GvHD assessment guidelines** for diagnosis (A) and severity scoring (B)

Legend supplementary Figure 5:
User self-reported comfort with GvHD assessment guidelines was categorized based on the pre-test survey question "How comfortable are you with using the above mentioned criteria in your daily practice on a Likert scale of 1-10 (1= Not at all comfortable; 10= extremely comfortable)": "low" comfort (response 4 or less), "moderate" comfort (response between 5 and 7), "high" comfort (response 8 or above).

A. GvHD Diagnosis

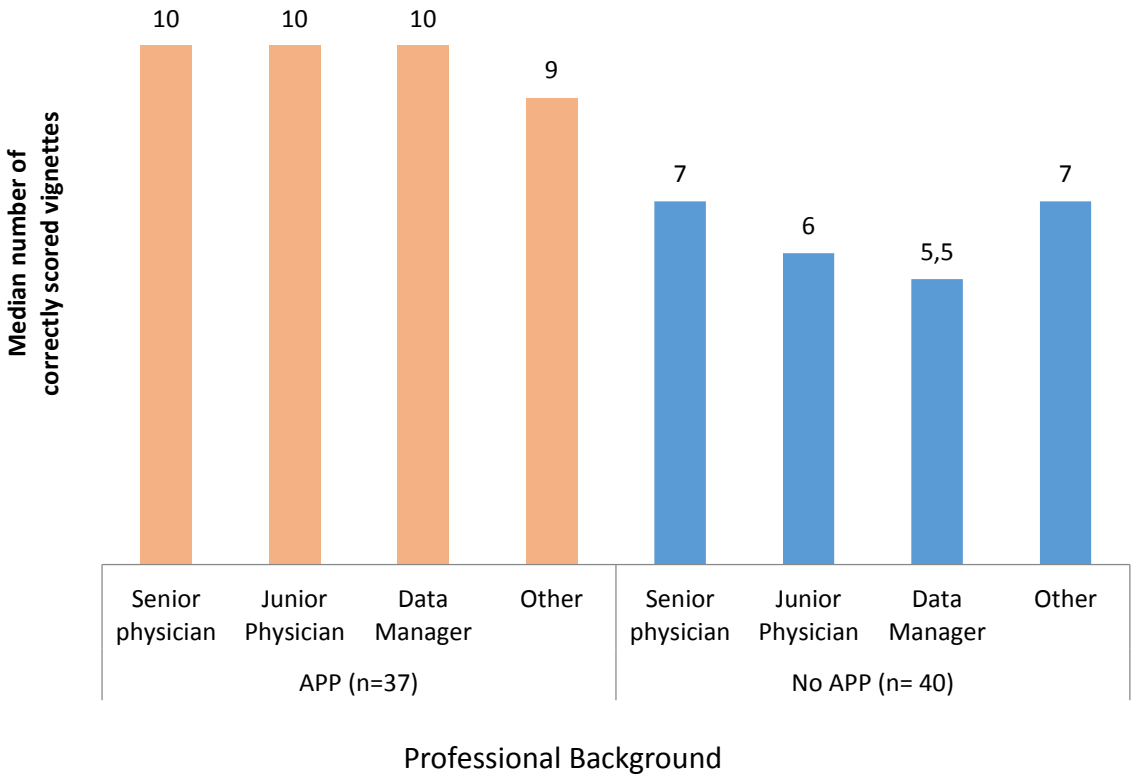


B. GvHD Severity Scoring



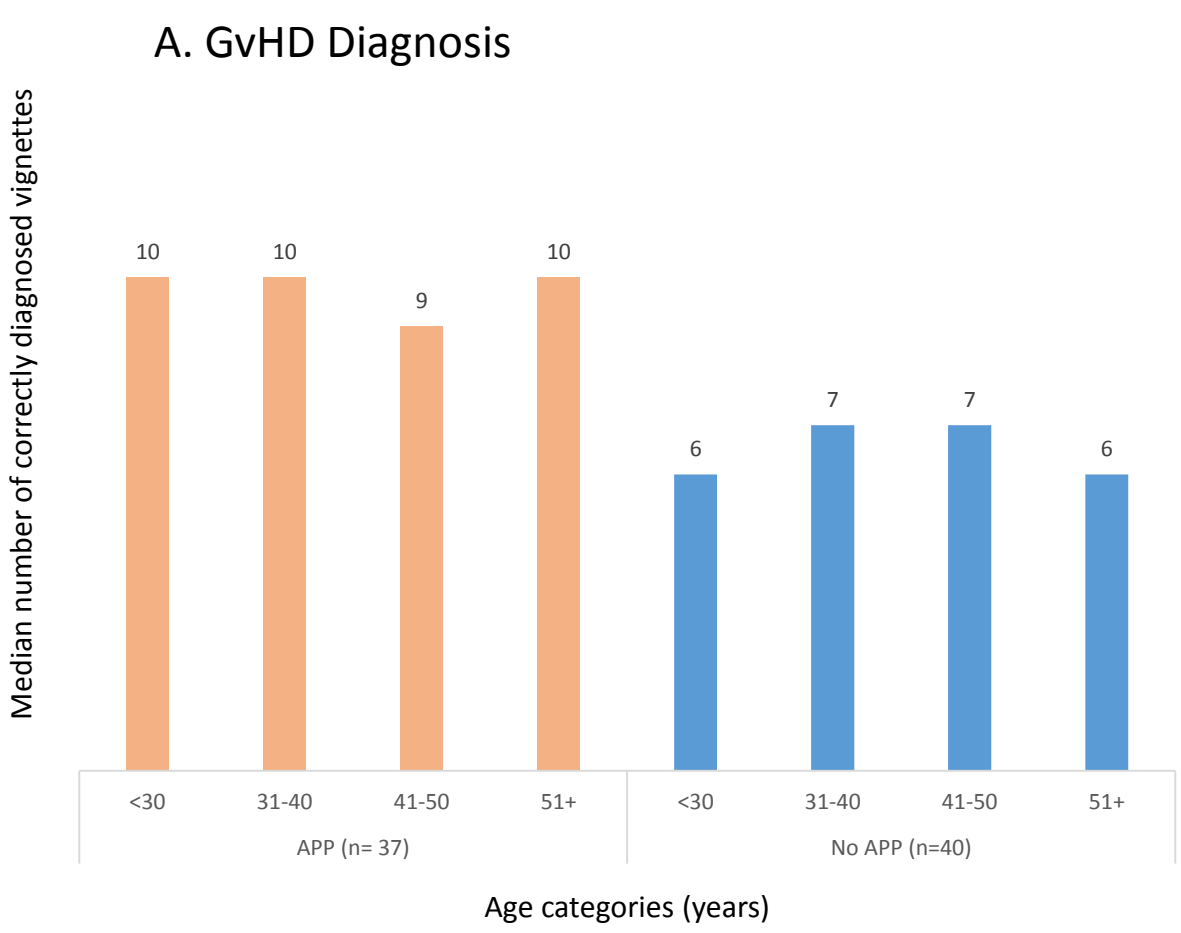
p=0.445

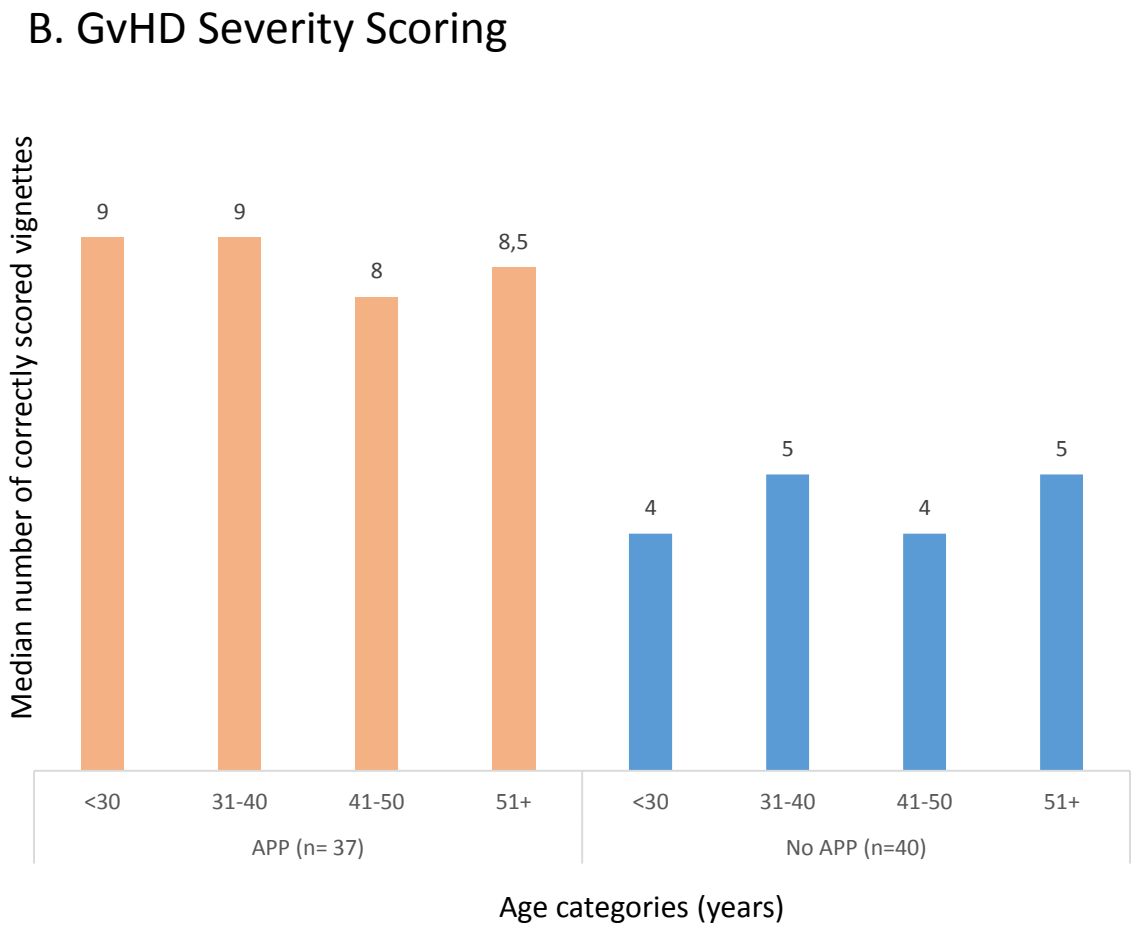
A. GvHD Diagnosis



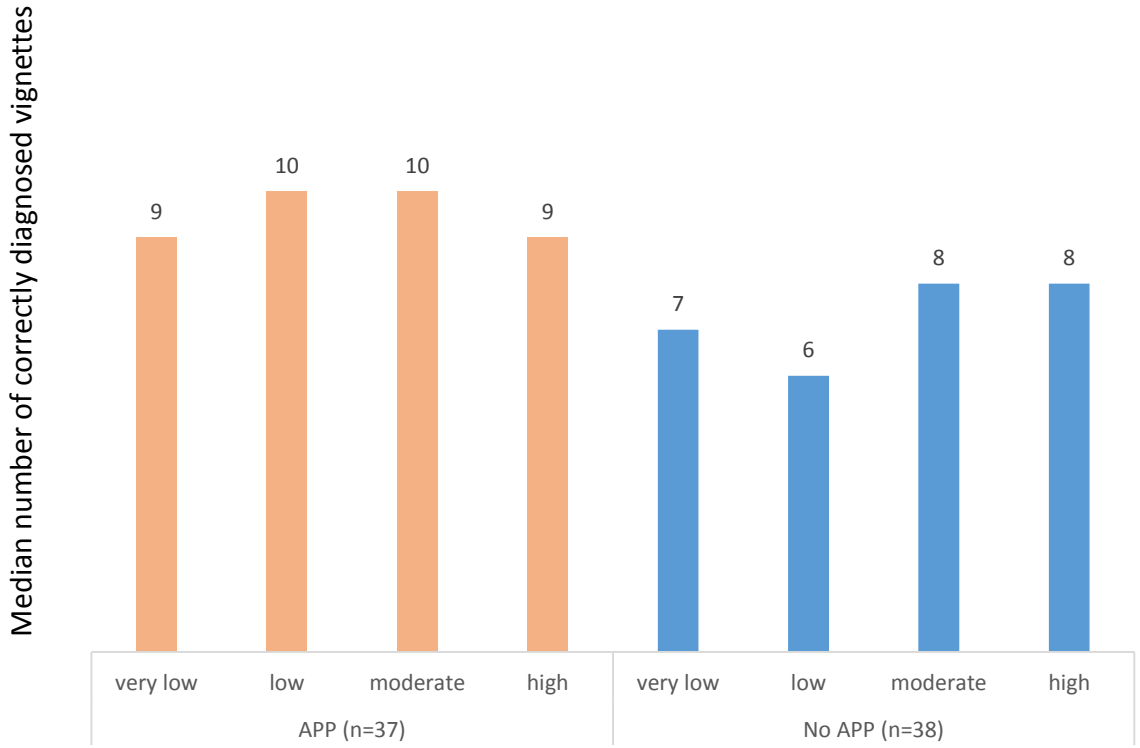
B. GvHD Severity Scoring

p=0.665

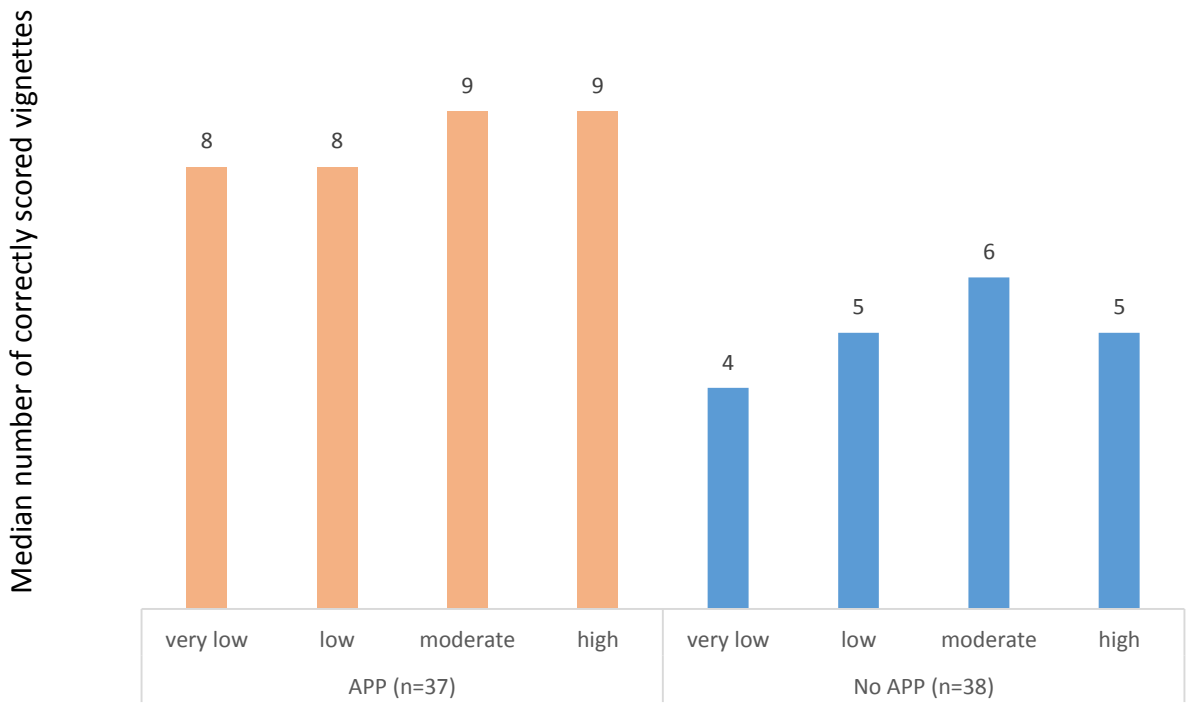

A. GvHD Diagnosis
Median number of correctly diagnosed vignettes
10
10
9
10
6
7
7
6
<30
31-40
41-50
51+
<30
31-40
41-50
51+
APP (n= 37)
No APP (n=40)
Age categories (years)


B. GvHD Severity Scoring
Median number of correctly scored vignettes
9
9
8
8,5
4
5
4
5
<30
31-40
41-50
51+
<30
31-40
41-50
51+
APP (n= 37)
No APP (n=40)
Age categories (years)

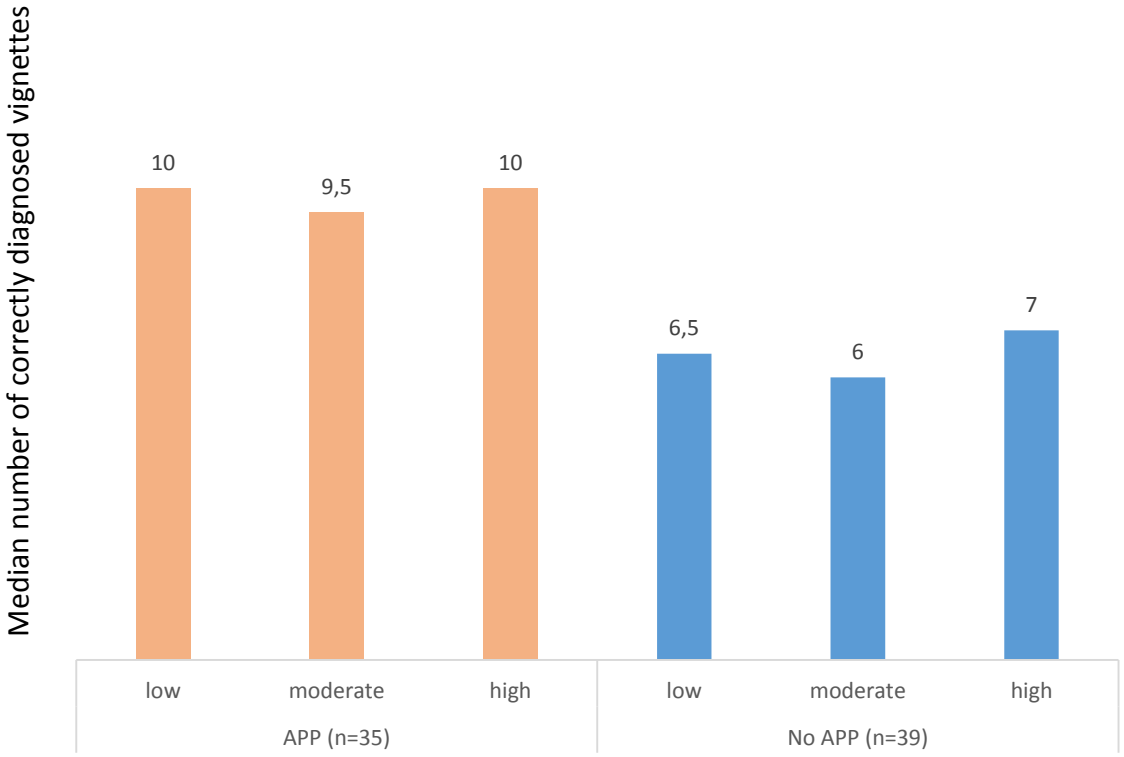
A. GvHD Diagnosis



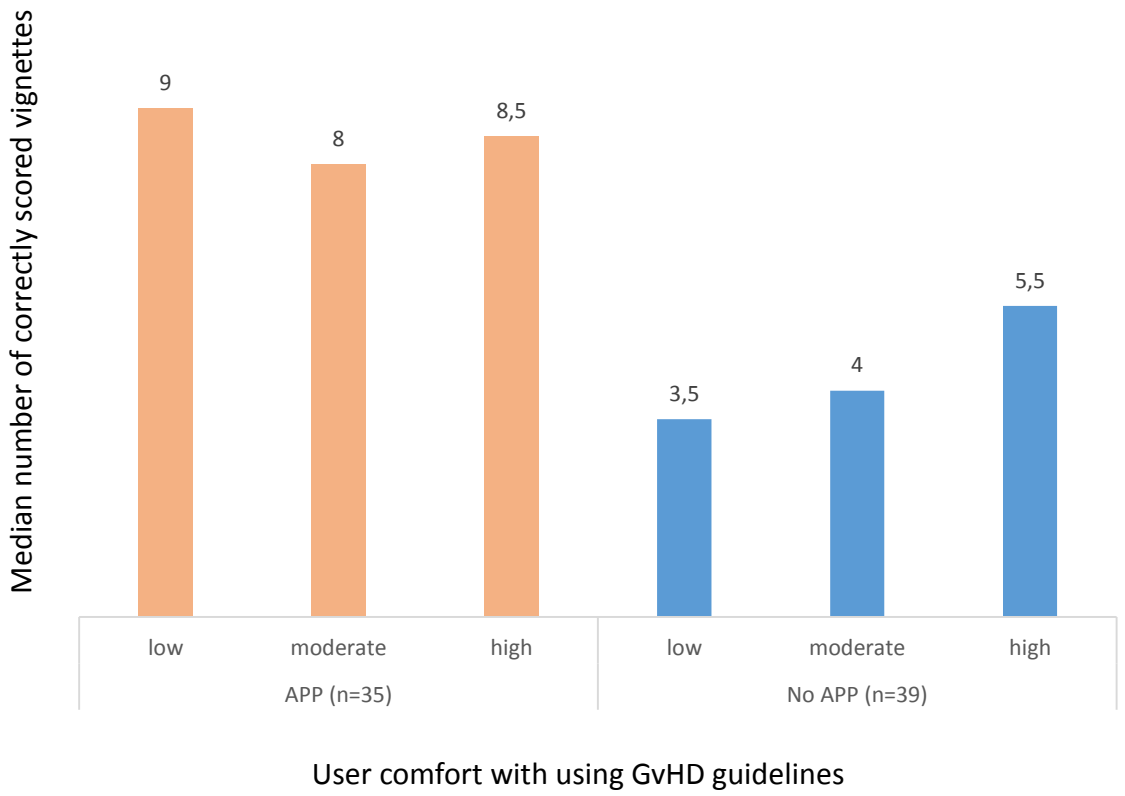
B. GvHD Severity Scoring



A. GvHD Diagnosis



B. GvHD Severity Scoring



Supplemental References

1. Harris AC, Young R, Devine S, et al. International, Multicenter Standardization of Acute Graft-versus-Host Disease Clinical Data Collection: A Report from the Mount Sinai Acute GVHD International Consortium. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2016;22(1):4-10.
2. Jagasia MH, Greinix HT, Arora M, et al. National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: I. The 2014 Diagnosis and Staging Working Group report. Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation. 2015;21(3):389-401 e381.
3. von Eye A, von Eye M. On the marginal dependency of cohen's κ. European Psychologist 2008;13(305-315).
4. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. Educational and Psychological Measurement 2016. 2016;76(609-637).