Dysphagia (2019) 34:308–314
https://doi.org/10.1007/s00455-018-9935-2

ORIGINAL ARTICLE

Validation of the German Version of the Yale Pharyngeal Residue Severity Rating Scale

Marco Gerschke[1] · Thomas Schöttker-Königer[2] · Annette Förster[1] · Jonka Friederike Netzebandt[3] · Ulla Marie Beushausen[2]

Received: 17 January 2018 / Accepted: 9 August 2018 / Published online: 16 August 2018

Abstract

The aim of this study was to validate the German Version of the Yale Pharyngeal Residue Severity Rating Scale and to investigate the impact of rater experience and training. The English original scale was translated into German language using a forward–backward method. For the validation, 30 images of different pharyngeal residue amounts have been selected and assigned to the scales' severity levels by two experts. 28 raters rated the images in randomized order in two passes. To investigate the impact of experience and training, raters were grouped into subgroups. Statistical analysis was carried out using kappa statistics. The results demonstrate excellent residue ratings for construct validity, inter-rater reliability and intra-rater reliability regarding overall group and all subgroups (Kappa > 0.90). No significant differences were found between groups by experience and inconstant differences were found between groups by training. The German version of the Yale Pharyngeal Residue Severity Rating Scale is a valid, reliable instrument for scoring location and severity of pharyngeal residue in the context of flexible endoscopic evaluations of swallowing.

Keywords Deglutition · Deglutition disorders · Pharyngeal residue · Rating scale · Flexible endoscopic evaluation of swallowing

Introduction

Pharyngeal residue of secretions or bolus material is a common symptom of swallowing disorders, caused by insufficient clearance in the pharyngeal swallowing stage [1]. Although being a surrogate parameter itself, pharyngeal residue is considered as a risk factor for post-swallow aspiration of secretions and food [1, 2]. Besides that, pharyngeal residue can affect the safety and effectiveness of medication intake and can lower patients' quality of life [3–5].

Fiberoptic endoscopic evaluation of swallowing (FEES) and videofluoroscopic swallow study (VFSS) are established methods in the field of diagnostics and management of dysphagia [6–8]. In Germany, FEES is the most widely used technique for stroke [9]. FEES training curricula and certifications are granted nationwide jointly by the German Society of Neurology/German Stroke Society/German Society of Geriatrics, Europewide by the European Society for Swallowing Disorders [10, 11]. With regard to identification of pharyngeal residue, FEES seems to be superior to VFSS but leads to more severe impressions [12, 13]. Severity rating of pharyngeal residue in FEES has been problematic as it is strongly related to the examiner's perception. Deviating residue assessments can result in different interpretations of findings, treatment strategies or dietary recommendations [14]. Multiple rating scales are available to assess pharyngeal residue on FEES. Only few of those have undergone validity and/or reliability testing

✉ Marco Gerschke
MGerschke@schoen-klinik.de

[1] Department of Neurology and Neurorehabilitation, Schön Klinik Hamburg Eilbek, Dehnhaide 120, 22081 Hamburg, Germany

[2] Faculty of Social Work and Health, HAWK University of Applied Sciences and Arts Hildesheim, Goschentor 1, 31134 Hildesheim, Germany

[3] Department of Speech- and Language Therapy, P.A.N. Center for Post-acute Neurorehabilitation, Wildkanzelweg 28, 13465 Berlin, Germany

[15]. While most residue rating scales are designed as ordinal scales, Pisegna et al. demonstrated high reliability for ratings using a visual analog scale [16, 17].

The Yale Pharyngeal Residue Severity Rating Scale is a validated, reliable tool to assess pharyngeal residue based on FEES [15, 18, 19]. Designed as a five-point ordinal scale, it allows subjective severity ratings (none, trace, mild, moderate, and severe) for residue located in the valleculae or the pyriform sinuses. Severity levels are defined by anatomical descriptions based on the image from an endoscopic perspective. The Yale Pharyngeal Residue Severity Rating Scale has shown very good to excellent psychometric properties with regard to construct validity, inter-rater reliability, and intra-rater reliability [18, 19]. To date, the Yale Pharyngeal Residue Severity Rating Scale has not been translated into other languages (P. Neubauer, personal communication, March 28, 2018).

For the translation of measuring instruments into foreign languages, it is recommended to follow a standardized translation process in order to ensure a suitable linguistic accuracy and cultural appropriateness. Furthermore, it is necessary to prove the measurement equivalence of the translated instrument [20, 21].

The aim of this study was to develop a German translation of the Yale Pharyngeal Residue Severity Rating Scale and to evaluate its construct validity, inter-rater reliability, and intra-rater reliability. In addition, the effects of rater experience and training were investigated.

Methods

Translation Process

The authorized translation process was based on the guidelines by Schmitt and Eid [20] and comprised the following steps:

- Forward translations
- Synthesis into a consensus version
- Backward translations
- Review of the backward translations
- Expert interviews
- Review of interview results

Two forward translations into German language were produced by two bilingual speech and language therapists with expertise in deglutology (both native German speakers with excellent English language skills). Both versions were compared, discussed, and merged into a consensus version by an expert committee, consisting of the translators and an external FEES expert. Two back translations of the consensus version into English language were carried out by two native English speakers with excellent German language skills, who were new to the instrument. The back translations were compared to the original English version and discussed by the expert committee. In addition, two expert interviews with FEES experts were carried out in order to investigate the understandability of the German version and its relevance to the German-speaking area. Interviews were transcripted and underwent content analysis [22]. In the final review stage, the interview results and the possible need for modifications were discussed by the expert committee.

Validation Process

The study design of this validation study was orientated closely on the study design of the scales' original version, based on hierarchical categorization of images using the German version of the scale [18, 19].

Image Selection

For the image selection process, 84 images from the original validation study were used. 83 of these images displayed bolus residue of yellow pudding or white milk from endoscopic view, one image displayed no residue. In order to enlarge the selection for the underrepresented severity level I (no residue), 11 further images with no residue were added from the portfolio of the first author, used for educational purposes. For lack of a valid gold standard in German language, referent values for residue severity ratings were determined by two FEES experts with a combined 27 years of expertise [23, p. 169]. Both experts are holders of the FEES trainer certificate [11]. In a first step of the selection process, the total of 95 images were categorized separately by both experts to the scales' severity levels. In order to obtain one reference value per image for the calculation of the construct validity, only images with consistent assessment by both experts were included (49 images); images with divergent rating were excluded (46 images). In a second step, 30 images (15 per location, three per severity level) were selected by consensus.

Raters

Raters were recruited from different institutions in Germany. The following inclusion criteria were based on the preconditions of the German FEES training curriculum: professional activity as a physician or speech and language therapist, regular completion of FEES with a minimum experience of 1 year [11].

Subgroups by FEES experience in years and training status were created. For the groups by experience, the total group was divided into two equal groups, based on median years of experience. For the composition of the groups by training, 50% of the raters from each group by experience were randomly selected to receive rater training prior to the data collection. The other 50% of the raters received no training. Randomization was carried using the software Research Randomizer [24].

Training was delivered though an eight-minute video tutorial in German language, including an audio-visual introduction to the scales' concept, depictions, and descriptions of the severity levels and rating exercises with images. No images of the later investigation were used for this purpose.

Data Collection

Ratings of the 30 randomized color images (15 vallecula, 15 piriform) were given via an online platform, and a second rating was carried out at an interval of 2 weeks in rerandomized order. Each image was displayed separately with a scaling of 720 × 476 pixels, including written information about the location to be classified and the corresponding scale. Ratings were made by clicking on the severity level. For best visibility, raters were asked to complete the task in full-screen mode.

Statistical Analysis

The extent of construct validity, intra-rater reliability, and inter-rater reliability were assessed using kappa statistics. *Construct validity* was calculated based on the agreement between the initial ratings and expert ratings using a pooled Fleiss kappa coefficient with quadratic weights. *Intra-rater reliability* was determined by the agreement of the first and second ratings calculating a pooled Fleiss–kappa with quadratic weights. For *inter-rater reliability*, the agreement of the initial ratings was calculated by the quadratic weighted Kappa coefficient. The benchmark system of Fleiss [23, p. 121] was used to interpret kappa values. Differences in kappa values between subgroups by years of experience and training were assessed using the paired sample *t* test. Gwet proposed the use of the paired sample *t* test for testing the difference between two correlated agreement coefficients for statistical significance [25]. With regard to our study, we had correlated samples as different raters judged the same subjects. Kappa values are reported ± their standard errors (se) and median years of experience ± interquartile range (IQR). Statistical analysis was carried out with Stata (StataCorp LLC, USA, Version 14.2). For the kappa statistics, the user written command kappaetc was used [26].

Ethics

No individuals underwent FEES in the context of this study. As all images were selected from pre-existing, non-identifiable material, no ethics board approval was necessary. In accordance with the Declaration of Helsinki of the World Medical Association, written informed consent was obtained from all raters prior to their participation.

Results

Results of the Translation Process

Both forward translations were almost identical. Minor adaptations regarding wording were made to produce the consensus version. Retranslations were largely consistent with the original English version, so no further adaptations were made for the pre-final version. The results of the expert interviews showed good overall understandability of the pre-final version and high relevance for the German-speaking area. The pre-final version was obtained as the final version with no further changes (Tables 1, 2).

Rater Characteristics

Rater characteristics are summarized in Table 3. A total of 28 raters participated in this study, 22 of whom were speech and language therapists (SLT) and 6 of whom were physicians. The median years of experience (± IQR) were 4.5 (± 8.5) for the overall group. Each of the 28 raters was assigned to groups by experience (1–4 years; ≥ 5 years) with a median experience of 2 years (± 3) and 10.5 (± 10). Untrained raters ($n = 14$) and trained raters

Table 1 English and German versions of severity definitions for vallecula residue

English version			
I	None	0%	No residue
II	Trace	1–5%	Trace coating of the mucosa
III	Mild	5–25%	Epiglottic ligament visible
IV	Moderate	25–50%	Epiglottic ligament covered
V	Severe	> 50%	Filled to epiglottic rim
German version			
I	Keine	0%	Keine Residuen
II	Spuren	1–5%	Spuren überziehen die Schleimhaut
III	Leicht	5–25%	Epiglottisches Ligament sichtbar
IV	Mäßig	25–50%	Epiglottisches Ligament bedeckt
V	Stark	> 50%	Gefüllt bis Epiglottisrand

Table 2 English and German version of severity definitions for pyriform sinus residue

English version			
I	None	0%	No residue
II	Trace	1–5%	Trace coating of the mucosa
III	Mild	5–25%	Up wall to quarter full
IV	Moderate	25–50%	Up wall to half full
V	Severe	>50%	Filled to aryepiglottic fold
German version			
I	Keine	0%	Keine Residuen
II	Spuren	1–5%	Spuren überziehen die Schleimhaut
III	Leicht	5–25%	Wandaufwärts bis zu einem Viertel gefüllt
IV	Mäßig	25–50%	Wandaufwärts bis zur Hälfte gefüllt
V	Stark	> 50%	Gefüllt bis zur aryepiglottischen Falte

(n = 14) had both a median of 4.5 years of experience (no training: ± 7; training: ± 9).

Results of the Validation

Results Across All Raters

Results show excellent kappa statistics (Kappa > 0.90) for both locations regarding construct validity, inter-rater validity, and intra-rater validity (Table 4).

Results by Years of Experience

Construct validity kappa statistics for both groups were between 0.942 (± 0.039) and 0.950 (± 0.030) for both locations. Inter-rater reliability kappa statistics ranged between 0.925 (± 0.030) and 0.940 (± 0.027) depending on location and group. Intra-rater reliability kappas for both groups and locations were between 0.939 (± 0.060) and 0.968 (± 0.023). We found no differences between groups by years of experience (Table 5).

Results by Training Status

Construct validity kappa statistics by training status were between 0.915 (± 0.059) and 0.979 (± 0.020) dependent upon residue location. There were no differences between both groups. Kappas for inter-rater reliability were between 0.915 (± 0.021) and 0.964 (± 0.013) for both locations. Trained raters had significantly higher kappa values for vallecula residue ratings (p = 0.007), but there was no difference for pyriform sinus kappas (p = 0.682). Intra-rater reliability kappa values ranged between 0.941 (± 0.040) and 0.952 (± 0.032) with no differences between trained and untrained raters (Table 6).

Table 4 Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics (standard error) for vallecula and pyriform sinus ratings across all raters

	Vallecula Kappa (se)	Pyriform sinus Kappa (se)
Construct validity	0.943 (± 0.033)	0.947 (± 0.025)
Inter-rater reliability	0.928 (± 0.026)	0.938 (± 0.027)
Intra-rater reliability	0.963 (± 0.028)	0.944 (± 0.051)

Based on 420 ratings

Discussion

In this study, the Yale Pharyngeal Residue Severity Rating Scale was translated from English into German language. The validation demonstrated excellent results regarding overall construct validity, inter-rater reliability, and intra-rater reliability for both locations (vallecula and pyriform sinus). No significant differences were found between the two groups of 1–4 and ≥ 5 years of experience regarding construct validity, inter-rater reliability, and intra-rater reliability, indicating that the scale is suitable for experienced and less-experienced clinicians. An uneven picture emerges for the influence of training: a significant difference in favor of trained raters was shown for inter-rater reliability for vallecula residue ratings. All other differences, however, did not turn out to be significant. In consideration of the fact that the determined kappa values of both groups were invariably higher 0.9, it can be stated that the German version of the YPRSRS is suitable for both trained and untrained raters. At the same time, the results indicate that the judgmental precision for vallecula residue

Table 3 Rater characteristics

	Total	Experience		Training	
		1–4 years	≥ 5 years	No training	Training
n (%)	28 (100%)	14 (50%)	14 (50%)	14 (50%)	14 (50%)
Profession n (%)					
SLT	22 (79%)	9 (64%)	13 (93%)	13 (93%)	9 (64%)
Physicians	6 (21%)	5 (36%)	1 (7%)	1 (7%)	5 (36%)
Median years of experience (IQR)	4.5 (8.5)	2 (3)	10.5 (10)	4.5 (7)	4.5 (9)

Table 5 Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics (standard error) for vallecula and pyriform sinus ratings by years of experience

	1–4 years (n = 14) Kappa (se)	≥ 5 years (n = 14) Kappa (se)	t test* p value
Construct validity			
Vallecula	0.944 (± 0.047)	0.948 (± 0.043)	0.956
Pyriform sinus	0.942 (± 0.039)	0.950 (± 0.030)	0.820
Inter-rater reliability			
Vallecula	0.925 (± 0.030)	0.928 (± 0.024)	0.866
Pyriform sinus	0.937 (± 0.027)	0.940 (± 0.027)	0.695
Intra-rater reliability			
Vallecula	0.958 (± 0.032)	0.968 (± 0.023)	0.782
Pyriform sinus	0.939 (± 0.060)	0.948 (± 0.039)	0.889

Based on 210 ratings

*Paired sample t test

Table 6 Construct validity, inter-rater reliability, and intra-rater reliability kappa statistics (standard error) for vallecula and pyriform sinus ratings by training versus no training

	No training (n = 14) Kappa (se)	Training (n = 14) Kappa (se)	t test* p value
Construct validity			
Vallecula	0.915 (± 0.059)	0.979 (± 0.020)	0.278
Pyriform sinus	0.940 (± 0.033)	0.951 (± 0.037)	0.776
Inter-rater reliability			
Vallecula	0.915 (± 0.021)	0.964 (± 0.013)	0.007
Pyriform sinus	0.933 (± 0.025)	0.938 (± 0.027)	0.682
Intra-rater reliability			
Vallecula	0.952 (± 0.032)	0.946 (± 0.060)	0.575
Pyriform sinus	0.941 (± 0.040)	0.943 (± 0.062)	0.948

Based on 210 ratings

*Paired sample t test

may be increased by training and rating experience. A set of exemplary images per location and severity level have been published with the original version of the Yale Pharyngeal Severity Rating Scale, which can be used as a Reference [18, 19].

In comparison to the original version, the German translation achieved similar results. In some areas, the results of the English version are even exceeded: Neubauer et al. [18] reported kappa values for inter-rater reliability of 0.868 for vallecula (0.928 in German) and 0.751 for pyriform sinus (0.938 in German). Intra-rater reliability for pyriform sinus was 0.854 in English (0.944 in German). The English version displayed inconsistent differences between groups by years of experience regarding inter-rater reliability for both locations [18]; in the German version, there were no significant differences. Between groups by training status, Neubauer et al. [18] found a significant difference for inter-rater reliability in favor of trained raters for both locations. This observation is also found in the results of the German version, albeit only for vallecula residue ratings, indicating that (i) vallecula residue might be more difficult to assess to the scales' severity levels and (ii) rating precision can be improved by training. Further, it can be assumed that the scale could have been known to some of the raters in advance to this study. The original version, published in 2015, was well received in German-speaking countries. It cannot therefore be ruled out that some raters already had previous experience with the English version.

Finally, some limitations should be mentioned. First, the translation process, despite comprising all essential steps, did not meet entirely international standards [21]. Pretesting took place as interviews with FEES experts followed by a review of the expert committee. Second, we used images for severity ratings. Hereby, we followed the original methodological approach to achieve best possible comparability. However, video sequences would have been closer to the examination situation and might have led to different results. Third, we used expert ratings as a 'construct.' In absence of a gold standard, we mirrored the approach of Neubauer et al. to calculate construct validity, which can be considered as a limitation [18, 19].

Furthermore, it has to be stated that image selection by consensus created a "best-of-the-best"-selection that led to the exclusion of less clear images. In real-life practice, the amount of residue can distribute unevenly in the pharyngeal cavities and a sufficiently clear view is not always given. Selecting images by consensus and not at random might have impacted the likelihood of agreement. Fourth, the depicted bolus residue consisted of milk or pudding. While these are typical foods used in FEES examinations, this selection represents only a portion of possible bolus types [7, p. 87]. Especially solid or less cohesive boluses may be more difficult to rate [16].

Conclusion

The validity and reliability of the German Version of the Yale Pharyngeal Residue Severity Rating Scale were confirmed. The tool allows accurate objectification of pharyngeal residue location and severity level in the context of FEES regardless of the clinician's experience. Rater training and use of the original reference images can be recommended to increase rating precision. Future studies should focus on using random images or video sequences and other bolus types to evaluate the psychometric properties under real-life conditions.

Acknowledgement The authors gratefully acknowledge Paul D. Neubauer, Yale School of Medicine for providing the original images and Daniel Klein, University of Kassel for his support and consultation with his Stata command "kappaetc."

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed Consent Written informed consent was obtained from all the participants before participation in this study.

References

1. Murray J, Langmore SE, Ginsberg S, Dostie A. The significance of accumulated oropharyngeal secretions and swallowing frequency in predicting aspiration. Dysphagia. 1996;11:99–103. https://doi.org/10.1007/BF00417898.
2. dos Santos RRD, Sales AVMN, Cola PC, et al. Association between pharyngeal residue and posterior oral spillage with penetration and aspiration in Stroke. CoDAS. 2014;26:231–4. https://doi.org/10.1590/2317-1782/201420140476.
3. Schiele JT, Penner H, Schneider H, et al. Swallowing tablets and capsules increases the risk of penetration and aspiration in patients with stroke-induced dysphagia. Dysphagia. 2015; 30:571–82. https://doi.org/10.1007/s00455-015-9639-9.
4. Leder SB, Lerner MZ. Nil per os except medications order in the dysphagic patient. QJM. 2013;106(1):71–5. https://doi.org/10.1093/qjmed/hcs044.
5. Meyer TK, Pisegna JM, Krisciunas GP, Pauloski BR, Langmore SE. Residue influences quality of life independently of penetration and aspiration in head and neck cancer survivors: residue affects both QoL and function. Laryngoscope. 2017; 127:1615–21. https://doi.org/10.1002/lary.26387.
6. Langmore SE, Kenneth SMA, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. Dysphagia. 1988;2:216–9. https://doi.org/10.1007/BF02414429.
7. Langmore S. Endoscopic evaluation and treatment of swallowing disorders. 2nd ed. New York: Thieme; 2001. ISBN 978-0-86577-838-2.
8. Schatz K, Langmore SE, Olson N. Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. Ann Otol Rhinol Laryngol. 1991;100:678–81. https://doi.org/10.1177/000348949110000815.
9. Suntrup S, Meisel A, Dziewas R, Ende F, Reichmann H, Heuschmann P, Ickenstein GW. Dysphagia diagnostics and therapy of acute stroke: federal survey of certified stroke units. Nervenarzt. 2012;83:1619–24. https://doi.org/10.1007/s00115-012-3611-9.
10. Dziewas R, Glahn J, Helfer C, et al. Flexible endoscopic evaluation of swallowing (FEES) for neurogenic dysphagia: training curriculum of the German Society of Neurology and the German stroke society. BMC Med Educ. 2016;16:70. https://doi.org/10.1186/s12909-016-0587-3.
11. Dziewas R, Baijens L, Schindler A, Verin E, Michou E, Clave P, The European Society for Swallowing Disorders. European society for swallowing disorders FEES accreditation program for neurogenic and geriatric oropharyngeal dysphagia. Dysphagia. 2017;32:725–33. https://doi.org/10.1007/s00455-017-9828-9.
12. Kelly AM, Leslie P, Beale T, Payten C, Drinnan MJ. Fibreoptic endoscopic evaluation of swallowing and videofluoroscopy: does examination type influence perception of pharyngeal residue severity. Clin Otolaryngol. 2006;31:425–32. https://doi.org/10.1111/j.1749-4486.2006.01292.x.
13. Pisegna JM, Langmore SE. Parameters of instrumental swallowing evaluations: describing a diagnostic dilemma. Dysphagia. 2016;31:462–72. https://doi.org/10.1007/s00455-016-9700-3.
14. Langmore SE. History of fiberoptic endoscopic evaluation of swallowing for evaluation and management of pharyngeal dysphagia: changes over the years. Dysphagia. 2017;32:27–38. https://doi.org/10.1007/s00455-016-9775-x.
15. Neubauer PD, Hersey DP, Leder SB. Pharyngeal residue severity rating scales based on fiberoptic endoscopic evaluation of swallowing: a systematic review. Dysphagia. 2016;31:352–9. https://doi.org/10.1007/s00455-015-9682-6.
16. Pisegna JM, Borders JC, Kaneoka A, Coster WJ, Leonard R, Langmore SE. Reliability of untrained and experienced raters on FEES: rating overall residue is a simple task. Dysphagia. 2018. https://doi.org/10.1007/s00455-018-9883-x.
17. Pisegna JM, Kaneoka A, Leonard R, Langmore SE. Rethinking residue: determining the perceptual continuum of residue on FEES to enable better measurement. Dysphagia. 2018;33(1): 100–8. https://doi.org/10.1007/s00455-017-9838-7.
18. Neubauer PD, Rademaker AW, Leder SB. The Yale pharyngeal residue severity rating scale: an anatomically defined and image-based tool. Dysphagia. 2015;30:521–8. https://doi.org/10.1007/s00455-015-9631-4.
19. Leder SB, Neubauer PD. The Yale pharyngeal residue severity rating scale. Cham: Springer; 2016. ISBN 978-3-319-29899-3.
20. Schmitt M, Eid M. Guidelines for the translation of foreign-language measurement instruments. Diagnostica. 2007;53:1–2. https://doi.org/10.1026/0012-1924.53.1.1.
21. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25:3186–91.

22. Mayring P. Qualitative inhaltsanalyse: grundlagen und techniken. 12th ed. Weinheim: Beltz; 2015. ISBN 978-3-407-25730-7.
23. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: a practical guide. New York: Cambridge University Press; 2011. ISBN 978-0-521-13385-2.
24. Urbaniak GC, Plous S. Research randomizer (Version 4.0) [computer software]; 2013. http://www.randomizer.org/.
25. Gwet KL. Testing the difference of correlated agreement coefficients for statistical significance. Educ Psychol Meas. 2016; 76:609–37. https://doi.org/10.1177/0013164415596420.
26. Klein D. KAPPAETC: Stata module to evaluate interrater agreement. Statistical Software Components S458283, Boston College Department of Economics; 2016. https://ideas.repec.org/c/boc/bocode/s458283.html.

Marco Gerschke MSc

Thomas Schöttker-Königer MSc

Annette Förster MD

Jonka Friederike Netzebandt MSc

Ulla Marie Beushausen PhD