




# Structured Reporting of Head and Neck Sonography Achieves Substantial Interrater Reliability



## Authors

Benjamin Philipp Ernst<sup>1</sup> , Carla Dörsching<sup>1</sup>, Alessandro Bozzato<sup>2</sup> , Jennis Gabrielpillai<sup>1</sup>, Sven Becker<sup>3</sup>, Matthias Frank Froelich<sup>4</sup> , Benedikt Kramer<sup>5</sup>, Christoph Sproll<sup>6</sup>, Mirco Schapher<sup>7</sup>, Miguel Goncalves<sup>8</sup>, Naglaa Mansour<sup>9</sup>, Benedikt Hofauer<sup>10</sup>, Wieland H Sommer<sup>11</sup>, Felix von Scotti<sup>12</sup>, Johannes Matthias Weimer<sup>13,#</sup>, Julian Künzel<sup>14,#</sup>

## Affiliations

- 1 Department of Otorhinolaryngology, University Medical Center Bonn, Bonn, Germany
- 2 Department of Otorhinolaryngology, Head & Neck Surgery, Saarland University Hospital and Saarland University Faculty of Medicine, Homburg, Germany
- 3 Department of Otorhinolaryngology, Head and Neck Surgery, University of Tübingen Medical Center, Tuebingen, Germany
- 4 Department of Radiology and Nuclear Medicine, University Medical Centre Mannheim, Mannheim, Germany
- 5 Department of Otorhinolaryngology, Head and Neck Surgery, University Medical Centre Mannheim, Mannheim, Germany
- 6 Department of Oral and Maxillofacial Surgery, Medical Faculty and University Hospital Düsseldorf, Duesseldorf, Germany
- 7 Department of Otorhinolaryngology, Head and Neck Surgery, Paracelsus Medical University, Nuremberg, Germany
- 8 Department of Otorhinolaryngology, Plastic Head and Neck Surgery, RWTH Aachen University Hospital, Aachen, Germany
- 9 Department of Otorhinolaryngology, University Medical Center Freiburg, Freiburg, Germany
- 10 Department of Otorhinolaryngology, Head and Neck Surgery, Technical University of Munich Hospital Rechts der Isar, Munich, Germany
- 11 Department of Radiology, LMU University Hospital, Munich, Germany
- 12 Ultrasound Division, Otorhinolaryngology Center Münsterland, Münster, Germany
- 13 Rudolf-Frey Teaching Department, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany
- 14 Department of Otorhinolaryngology, Universitätsklinikum Regensburg, Regensburg, Germany

## Key words

ultrasonography, structured reporting, quality improvement, interrater reliability

received 30.07.2023

accepted after revision 14.08.2023

## Bibliography

Ultrasound Int Open 2023; 9: E26–E32

DOI 10.1055/a-2173-3966

ISSN 2199-7152

© 2023. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

## Correspondence

Dr. Benjamin Philipp Ernst  
University Medical Center Bonn, Department of Otorhinolaryngology, Venusberg-Campus 1  
53127 Bonn  
Germany  
Tel.: +4922828713705  
Benjamin.Ernst@ukbonn.de

## ABSTRACT

**Purpose** Ultrasound examinations are often criticized for having higher examiner dependency compared to other imaging techniques. Compared to free-text reporting, structured reporting (SR) of head and neck sonography (HNS) achieves superior time efficiency as well as report quality. However, there are no findings concerning the influence of SR on the interrater reliability (IRR) of HNS.

# These authors contributed equally to this work and share senior authorship

**Materials and Methods** Typical pathologies (n = 4) in HNS were documented by video/images by two certified head and neck ultrasound instructors. Consequently, structured reports of these videos/images were created by n = 9 senior physicians at departments of otolaryngology or maxillofacial surgery with DEGUM instructors on staff. Reports (n = 36) were evaluated regarding overall completeness and IRR. Additionally, user satisfaction was assessed by a visual analog scale (VAS).

**Results** SR yielded very high report completeness (91.8%) in all four cases with a substantial IRR (Fleiss'  $\kappa$  0.73). Interrater agreement was high at 87.2% with very good user satisfaction (VAS 8.6).

**Conclusion** SR has the potential to ensure high-quality examination reports with substantial comparability and very high user satisfaction. Furthermore, big data collection and analysis are facilitated by SR. Therefore, process quality, workflow, and scientific output are potentially enhanced by SR.

## Background

Head and neck sonography (HNS) can be considered the diagnostic method of first choice for the diagnosis of a wide array of soft-tissue diseases, both in otolaryngology and maxillofacial surgery [1–3]. This is first and foremost based on its broad availability, patient safety, as well as the potential for intraoperative use [4]. Additionally, unlike computed tomography (CT) and magnetic resonance imaging (MRI), ultrasound is associated with lower costs and – unlike CT – is not associated with ionizing radiation and it is well suitable for patients suffering from claustrophobia [5]. Ultrasound imaging has undergone sustainable technological improvements in the past decades with vastly improving image quality and dynamic range.

Despite these advantages, ultrasound examinations have traditionally been associated with a significantly higher examiner dependency than CT and MRI examinations [5]. In the absence of data concerning the interrater reliability (IRR) and interrater agreement (IRA) of ultrasound as well as CT and MRI studies of the neck, existing data show a great variability of these parameters in each modality with respect to a specific disease and body region [6]. This may pose a major problem, especially in the preoperative workup of soft-tissue pathologies in head and neck surgery, since insufficient preoperative ultrasound reports may lead to major intraoperative complications, extension of surgery, or reoperations [7]. Education standards concerning ultrasound vary greatly among countries and also among medical schools within a country [8]. Consequently, the European Federation of Societies for Ultrasound in Medicine and Biology has proposed standards for HNS, both in terms of training and clinical practice [9, 10]. Until recently, there was no uniform standard regarding the structure and content of HNS reports in German-speaking countries [11]. Therefore, it is not surprising that the overall quality of HNS reports within a sample of German university medical centers has great potential for optimization [12].

Even despite harmonization of training and clinical practice, the mode of reporting has been pointed out as a major contributor to information loss and dissatisfaction among referring physicians [13]. Several studies were able to demonstrate that structured reporting (SR) improves the report quality and time efficiency of HNS by standardizing its content for various educational levels [8, 13–15]. Furthermore, SR can be considered a valuable tool to improve preoperative evaluation of CT scans in the context of functional endoscopic sinus surgery, both by radiologists and otolaryngologists

[16, 17]. Due to the standardized structure, SR is associated with a decreased likelihood of missing or misinterpreting key structures, potentially resulting in misdiagnosis. Additionally, previous studies were able to show a preference for SR by surgeons [18]. Considering these findings, SR might also improve IRR and IRA which may further promote the value of HNS in otolaryngology and maxillofacial surgery.

Therefore, the present study was designed to assess the impact of SR on the IRR of HNS in a cohort of experienced HNS examiners.

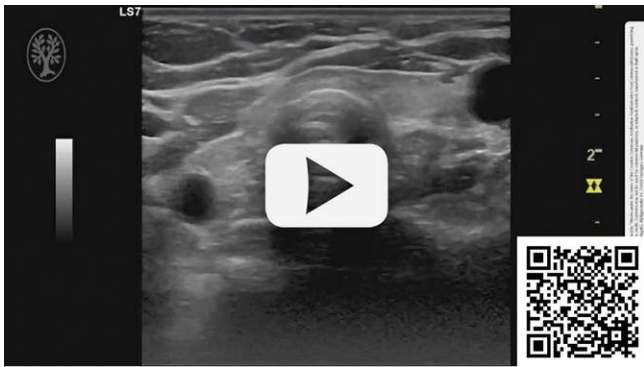
## Methods

### Study design

Video sequences of four complete HNS examinations as well as detailed images of the respective pathologies were recorded by two certified HNS instructors. An additional HNS examination of a healthy volunteer without pathological findings was used for training purposes (see **Video. 1**). The four cases included a follow-up for carcinoma of the parotid gland (case 1, see **Video. 2**), an evaluation of a parotid gland mass (case 2, see **Video. 3**), an evaluation of a suspected cervical lymph node metastasis (case 3, see **Video. 4**), as well as another evaluation of a parotid gland mass (case 4, see **Video. 5**). For these cases, the instructors created SRs using an online-based reporting template for HNS (Smart Reporting GmbH, Munich, Germany, <http://www.smart-reporting.com>), which were used as reference reports [8, 13–15]. Anonymized video and image files as well as detailed instructions on how to use the SR template were sent out to nine departments with significant HNS expertise.



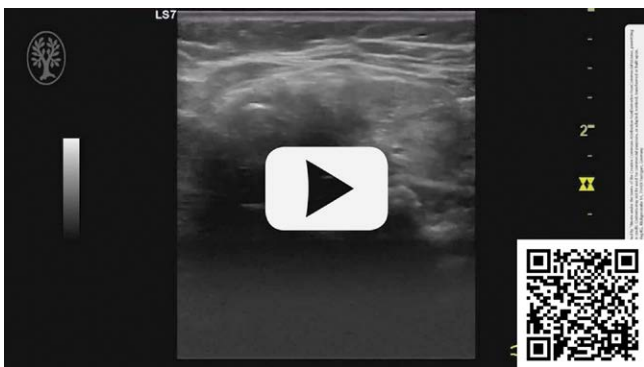
▶ **Video. 1** Video and image files of test case.



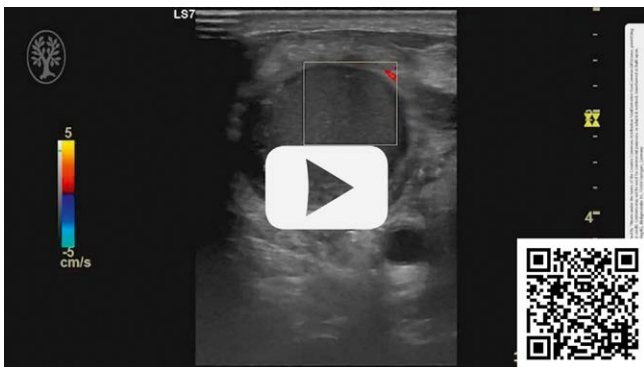
► **Video. 2** Video and image files of case 1: Follow-up for carcinoma of the parotid gland.



► **Video. 5** Video and image files of case 4: Evaluation of parotid gland mass.



► **Video. 3** Video and image files of case 2: Evaluation of parotid gland mass.



► **Video. 4** Video and image files of case 3: Evaluation of suspected cervical lymph node metastasis.

As previously described, the additional training case was included to allow the examiners to become familiar with the SR template and was to be completed first. Subsequently, participating senior physicians were asked to create SRs of the four cases based on the provided video and image files, resulting in  $n = 36$  reports. Since previous publications by our group showed that SR is consistently superior to conventional reporting (CR) in terms of report quality, there is sufficient data on the IRR of CR, and CR has limited relevance to the central aim of this study, no control group in which CR was used was included in the study design [8, 13–15].

After completing the SRs, the examiners rated the user friendliness of the SR template by using an existing questionnaire with a visual analog scale (VAS) [8, 13–15].

### Report evaluation

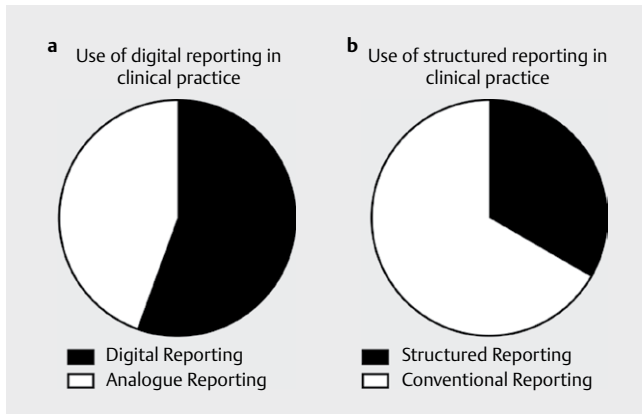
Two certified head and neck ultrasound instructors analyzed all 36 anonymized reports for overall report completeness as well as report content and assessed the IRR and IRA. In this scenario, IRR refers to the extent to which different examiners provide consistent evaluations, whereas IRA refers to the degree to which different examiners agree upon the same categorical decision or classification when rating the same content.

Participating senior physicians were questioned about user satisfaction utilizing an existing questionnaire [8, 13–15]. This questionnaire surveyed whether an SR template is useful and applicable in everyday clinical practice (questions 1 and 2) as well as whether SR may improve overall reporting (question 3). It also asked participating physicians about the time required for SR (question 4) and its economic value (question 5). Using a 10-point visual analog scale (VAS), the questionnaire furthermore asked whether SR might assist inexperienced physicians in learning ultrasound examinations and reporting (questions 6 and 7) and whether the SR template is easy to use and is neatly arranged (questions 8 and 9).

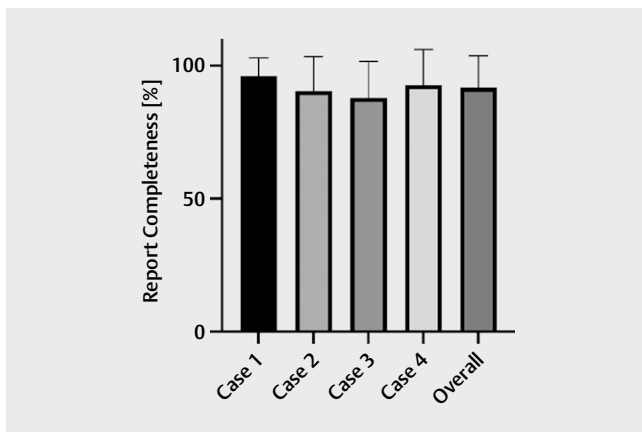
### Sample size calculation and statistical analysis

As described by Sim and Wright, the number of reports needed in this study was determined based on previous studies concerning SR of HNS [13]. The power was set to 80% with a significance level of  $\alpha = 0.05$ . Taking into account a proportion of positive ratings of 50%, a baseline  $\kappa$  of 0.4 and a previously published  $\kappa$  using SR of 0.9, 27 ratings are needed to determine significant differences in IRR [13]. The  $\kappa$  values were interpreted as proposed by the Landis and Koch classification [19]. Consequently, IRR was considered as almost perfect ( $\kappa$  0.81–1.0), substantial ( $\kappa$  0.61–0.8), moderate ( $\kappa$  0.41–0.6), adequate ( $\kappa$  0.21–0.4), or slight ( $\kappa$  0–0.2).

Data are reported as the mean  $\pm$  SD. The Shapiro-Wilk test was used to determine normal distributions. A T-test was used to compare overall completeness and IRA. A p-value of less than 0.05 was considered statistically significant. Fleiss'  $\kappa$  was used to evaluate IRR. All statistical analyses were performed using GraphPad Prism



► **Fig. 1** Distribution of reporting system use by participating senior physicians in the clinical routine. Within this cohort, 55.6% of participants used digital reporting systems in clinical practice (a) while 33.3% employed structured reporting elements (b).



► **Fig. 2** Results of overall report completeness analysis. Structured reporting (SR) achieves very high ratings in terms of report quality consistently throughout all four cases. No significant differences were observed among cases.

9.0.1 (Graphpad Software LLC., San Diego, CA, USA) and Microsoft Excel 2019 (Microsoft Corporation, Redmond, WA, USA).

## Results

Five of the nine participating departments (55.6%) reported use of a digital reporting system in clinical practice. Out of these, three departments (33.3%) used some kind of structured reporting approach (see ► **Fig. 1**).

### Report Completeness

In-depth analysis of reports created by study participants using SR revealed very high completeness ratings of all reports (91.8% ± 11.72%), which was consistent in all four cases (see ► **Fig. 2**). In detail, overall report completeness was 96.1% ± 6.5% for case 1 (follow-up after a parotid gland carcinoma), 90.5% ± 12.1% for case 2 (parotid gland mass), 87.9% ± 12.9% for case 3 (cervical mass) and 92.8% ± 12.6% for case 4 (parotid gland mass). Differences be-

tween the overall completeness ratings of the four cases were not significant.

### Interrater reliability and interrater agreement

The IRR was calculated using Fleiss'  $\kappa$  for each case as well as for all acquired data. Overall, the IRR was substantial with a Fleiss'  $\kappa$  of 0.73. Overall, the IRA was high at 87.2% ± 15.1%. In detail, Fleiss'  $\kappa$  was 0.78 and the IRA was 87.2% ± 15.1% for case 1, 0.92 and 96.9% ± 5.4% for case 2, 0.66 and 80.4% ± 13.1% for case 3, and 0.74 and 85.7% ± 18.2% for case 4. There was a significant difference in IRA between cases 2 and 3 ( $p=0.0177$ , see ► **Fig. 3**), while the other cases did not show significant differences.

### User satisfaction

Assessment of VAS-based questionnaires revealed very high user satisfaction using the SR template (8.6 ± 1.8). In detail, the SR-based approach was rated to be useful (9.8 ± 0.6) and suitable for routine clinical use (9.9 ± 0.3). SR was thought to improve reporting (9.6 ± 0.8) and to be time-efficient (7.2 ± 2) and participants felt that any additional time needed was well-spent (9 ± 0.9). Participating senior physicians stated that SR may be beneficial for inexperienced physicians to acquire ultrasound examination (8.2 ± 2.6) and reporting skills (9 ± 1.3). The template was perceived as easy to use (7.7 ± 0.9) and neatly arranged (7.3 ± 2.3, see ► **Fig. 4**).

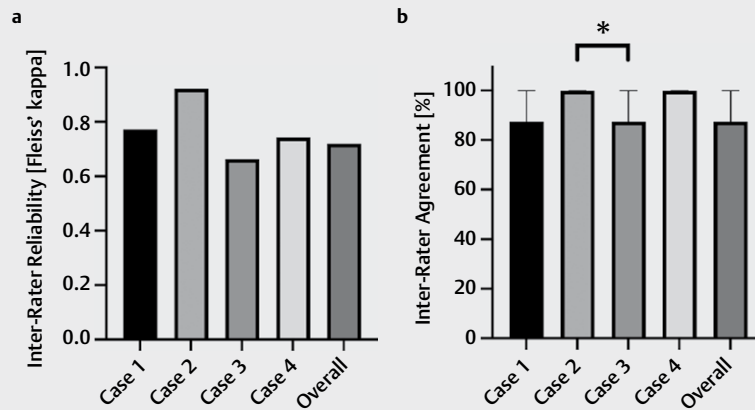
## Discussion

HNS is considered the diagnostic modality of choice for a wide variety of soft-tissue pathologies of the head and neck. To date, the preferable reporting modality has not yet been well defined [1–3], but SR seems to additionally increase the value of HNS examinations, as shown in different studies over the last couple of years [8, 13–15]. While report quality and the reliability of extracted information by referring physicians have been shown to be very high, there is, to our knowledge, no data concerning the impact of SR on the IRR of HNS. Consequently, the present study was designed to assess the IRR of HNS for various pathologies.

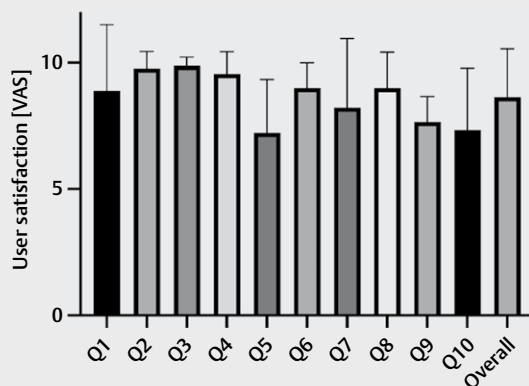
IRR is traditionally believed to be rather low for all kinds of ultrasound examinations compared to CT or MRI [5]. Especially for diagnostic modalities which are used for clinical follow-ups and may involve various examiners, IRR is of utmost importance as it reduces both false-positive and false-negative findings. While false-positive findings may trigger additional and unneeded, possibly cost-intensive or invasive diagnostic procedures and treatments, false-negative findings may lead to delayed diagnosis with progression of the underlying disease, potential decrease in prognosis, and possible legal consequences.

Therefore, the present study's results are of great interest as they underline the value of HNS as a quick, cost-efficient, noninvasive, and precise diagnostic modality. The encouraging findings of Goncalves et al. regarding the very good IRR of HNS for the assessment of sialolithiasis [5] which may be superior to CT or MRI for this indication are also extremely interesting.

Our data show that within this cohort report completeness using SR was very high for all four different cases. As shown in previous studies, implementation of SR improves report completeness especially through the standardized query of structures and regions,



► **Fig. 3** Results of interrater reliability and interrater agreement analysis. The use of structured reporting yields substantial to almost perfect inter-rater reliability in all analyzed cases (a). Additionally, interrater agreement was also very high in all cases (b). Except for cases 2 and 3, there were no significant differences in interrater agreement between analyzed cases. \* $p < 0.05$



► **Fig. 4** Analysis of questionnaire findings using visual analog scale (VAS, 0: complete disagreement, 10: complete agreement). Participating senior physicians were surveyed concerning the usefulness (Q1) and applicability (Q2) of structured reporting (SR) in everyday clinical practice, improvement in overall reporting (Q3), time efficiency (Q4), whether additionally needed time was well spent using SR (Q5), whether SR is beneficial for inexperienced physicians to learn ultrasound examinations (Q6) and reporting (Q7), whether the SR template is easy to use (Q8), and whether the SR template is neatly arranged (Q9). In summary, the questionnaire revealed substantial user satisfaction in all categories (overall).

even if they are not involved in the pathology of interest and are not in the center of attention [8, 13–15]. This is safeguarded by the appropriate use of mandatory items within the report template. Reporting may only proceed once these mandatory items are completed. This basic principle of SR has been proven to reduce the frequency of missed pathologic findings and to improve diagnostic precision [20].

Secondly, SRs utilize standardized terminology that has been previously approved in expert consensus and in accordance with published recommendations [11]. This ensures objective description of pathological findings. Moreover, the standardized terminol-

ogy, structure, and digitalization of SRs enable appropriate comparability of reports and scientific use in individual and big data analyses as well as the application of artificial intelligence and deep learning technologies [21].

Unlike CT and MRI, ultrasound is a dynamic examination technique, including the movement of structures and images, various angles, compressibility, and functional parameters such as the Doppler effect. Consequently, this entails a greater dependency on the individual examiner. Since participating physicians did not perform the ultrasound examinations on their own, standardized video sequences as well as detailed images depicting all necessary aspects of the pathology were provided in order to assess the IRR and IRA of HNS in a realistic manner. In addition to the very high completeness ratings, the IRR was substantial with a Fleiss'  $\kappa$  of 0.73. There were no significant differences in completeness or IRR for the different cases. The IRA was consistently very high ( $87.2\% \pm 15.1\%$ ) except for in two cases. Likewise, the standardized structure and terminology are major factors contributing to the substantial IRR.

Our data clearly demonstrate that the use of SR resulted in a consistent interpretation of the provided examination data. This is essential for a reliable diagnosis and efficient therapy. Due to the superior comparability, SR has the potential to improve communication with other involved healthcare providers, thus facilitating patient management and reducing inquiries and, more importantly, misunderstandings [22]. Furthermore, SR may be a valuable quality control tool to assist in the accreditation process that forms the basis for patient referrals, treatment, and billing in many countries [23].

Potential linguistic imperfections in written findings by non-native speakers are exacerbated by the continued increase in the demand for telemedicine solutions [24]. Telemedicine, which is particularly useful for reporting diagnostic modalities, has become a necessity for rural regions with a shortage of specialists [25]. The increasing availability of broadband internet connections has made it possible to transfer very large amounts of medical data that can be processed in other regions, whether nationally or internationally. In the case of international telemedicine reporting, there is a

risk that the reporting specialist may not have adequate linguistic competence to report findings or to respond to follow-up questions from referrers in the language of the source country. This may further hinder IRR in the context of HNS. Consequently, SR could be an essential element to overcome inadequate reporting quality due to poor language skills, as modern SR systems can automatically output a native language report in foreign languages [26].

Examiner satisfaction was very high in all ten assessed categories, with an overall VAS value of  $8.6 \pm 1.8$ . Our findings are in line with the literature and confirm the importance of SR for the quality of ultrasound examinations [5, 8, 13–15] but also in the context of big data analysis and therapy monitoring [18, 27]. Both the redundancy of the SR process and the standardized workflow are major contributors to the examiners' preference for SR compared to CR [8, 13–15]. Concerns on the part of physicians using analog CR that SR templates have overly rigid reporting conditions for the great variety of pathological findings in clinical practice have been rebutted by multiple studies [8, 13, 14, 28]. In fact, the opposite seems to be the case, since SR's rather rigid approach has been shown to be rather convenient for inexperienced examiners [8, 14].

## Conclusion

Our data demonstrate that the implementation of SR ensures a substantial IRR of HNS examinations, thereby reducing one of its most criticized disadvantages. The use of SR in clinical practice can improve diagnostic accuracy and safety of treatment, as well as simplify data analysis and transfer, communication, and quality assurance. Further studies will have to determine the potential impact on patient outcomes.

## Conflict of Interest

Wieland H Sommer is the founder of the company Smart Reporting GmbH, which hosts an online platform for structured reporting. The other authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article. This manuscript is part of medical doctoral theses presented by Carla Dörsching at the Medical Faculty of the University of Bonn.

## Acknowledgement

We would like to thank Mrs. Celine Christe for her help in revising the figures and videos.

## References

- [1] Coltrera MD. Clinician-performed thyroid ultrasound. *Otolaryngol Clin North Am* 2014; 47: 491–507. DOI: 10.1016/j.otc.2014.04.001
- [2] Bialek EJ, Jakubowski W, Zajkowski P et al. US of the major salivary glands: anatomy and spatial relationships, pathologic conditions, and pitfalls. *Radiographics* 2006; 26: 745–763. DOI: 10.1148/rg.263055024
- [3] Künzel J, Bozzato A, Strieth S. Sonographie in der Nachsorge bei Kopf- und Halskarzinomen. [Follow-up ultrasound of head and neck cancer]. *HNO* 2017; 65: 939–952. DOI: 10.1007/s00106-017-0411-3
- [4] Beggs AD, Thomas PR. Point of use ultrasound by general surgeons: review of the literature and suggestions for future practice. *Int J Surg* 2013; 11: 12–17. DOI: 10.1016/j.ijsu.2012.11.014
- [5] Goncalves M, Mantsopoulos K, Schapher ML et al. Interrater Reliability of Ultrasound in the Diagnosis of Sialolithiasis. *Ultraschall Med* 2019; 40: 481–487. DOI: 10.1055/a-0837-0712
- [6] Takahashi Y, Fujino Y, Miura K et al. Intra- and inter-rater reliability of rectus femoris muscle thickness measured using ultrasonography in healthy individuals. *Ultrasound J* 2021; 13: 21. DOI: 10.1186/s13089-021-00224-8
- [7] Cohen O, Blank A, Meiersdorf S et al. Impact of high-quality ultrasound following community ultrasound on surgical planning and active surveillance in patients with thyroid cancer. *Clin Endocrinol (Oxf)* 2021; 94: 990–997. DOI: 10.1111/cen.14415
- [8] Ernst BP, Katzer F, Künzel J et al. Impact of structured reporting on developing head and neck ultrasound skills. *BMC Med Educ* 2019; 19: 102. DOI: 10.1186/s12909-019-1538-6
- [9] Todsén T, Ewertsen C, Jenssen C et al. Head and Neck Ultrasound - EFSUMB Training Recommendations for the Practice of Medical Ultrasound in Europe. *Ultrasound Int Open* 2022; 8: E29–E34. DOI: 10.1055/a-1922-6778
- [10] Wustner M, Radzina M, Calliada F et al. Professional Standards in Medical Ultrasound - EFSUMB Position Paper (Short Version) - General Aspects. *Ultraschall Med* 2022; 43: 456–463. DOI: 10.1055/a-1854-2936
- [11] Kunzel J, Bozzato V, Ernst BP et al. Musterbefund zur sonografischen Diagnostik im Kopf-Hals-Bereich (Sektion Kopf-Hals der DEGUM.) [Standard reporting for head and neck ultrasound - a proposal]. *Ultraschall Med* 2022. DOI: 10.1055/a-1810-7173
- [12] Kunzel J, Bozzato A, Ernst BP et al. Qualität in der Befundung von Kopf- und Halssonographien an Universitätskliniken – eine Stichprobe [Quality in the appraisal of head and neck sonography results in university hospitals-a random sample]. *HNO* 2021; 69: 907–912. DOI: 10.1007/s00106-020-00989-9
- [13] Ernst BP, Hodeib M, Strieth S et al. Structured reporting of head and neck ultrasound examinations. *BMC Med Imaging* 2019; 19: 25. DOI: 10.1186/s12880-019-0325-5
- [14] Ernst BP, Strieth S, Katzer F et al. The use of structured reporting of head and neck ultrasound ensures time-efficiency and report quality during residency. *Eur Arch Otorhinolaryngol* 2020; 277: 269–276. DOI: 10.1007/s00405-019-05679-z
- [15] Ernst BP, Strieth S, Künzel J et al. Evaluation of optimal education level to implement structured reporting into ultrasound training. *Med Ultrason* 2020. DOI: 10.11152/mu-2530
- [16] Ernst BP, Reissig MR, Strieth S et al. The role of structured reporting and structured operation planning in functional endoscopic sinus surgery. *PLoS One* 2020; 15: e0242804. DOI: 10.1371/journal.pone.0242804
- [17] Becker S, Gonser P, Haas M et al. ENT Residents Benefit from a Structured Operation Planning Approach in the Training of Functional Endoscopic Sinus Surgery. *Medicina (Kaunas)* 2021; 57: DOI: 10.3390/medicina57101062
- [18] Norenberg D, Sommer WH, Thasler W et al. Structured Reporting of Rectal Magnetic Resonance Imaging in Suspected Primary Rectal Cancer: Potential Benefits for Surgical Planning and Interdisciplinary Communication. *Invest Radiol* 2017; 52: 232–239. DOI: 10.1097/RLI.0000000000000336

- [19] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174. DOI: 10.2307/2529310
- [20] Tuncyurek O, Garces-Descovich A, Jaramillo-Cardoso A et al. Structured versus narrative reporting of pelvic MRI in perianal fistulizing disease: impact on clarity, completeness, and surgical planning. *Abdom Radiol (NY)* 2019; 44: 811–820. DOI: 10.1007/s00261-018-1858-8
- [21] Shen YT, Yue WW, Xu HX. Editorial: Ultrasound in Oncology: Application of Big Data and Artificial Intelligence. *Front Oncol* 2021; 11: 819487. DOI: 10.3389/fonc.2021.819487
- [22] Sluijter CE, van Lonkhuijzen LR, van Slooten HJ et al. The effects of implementing synoptic pathology reporting in cancer diagnosis: a systematic review. *Virchows Arch* 2016; 468: 639–649. DOI: 10.1007/s00428-016-1935-8
- [23] Johannesen DTS, Wiig S. Exploring hospital certification processes from the certification body's perspective - a qualitative study. *BMC Health Serv Res* 2020; 20: 242. DOI: 10.1186/s12913-020-05093-w
- [24] Matsumoto M, Koike S, Kashima S et al. Geographic Distribution of Radiologists and Utilization of Teleradiology in Japan: A Longitudinal Analysis Based on National Census Data. *PLoS One* 2015; 10: e0139723. DOI: 10.1371/journal.pone.0139723
- [25] Johnston K, Smith D, Preston R et al. “From the technology came the idea”: safe implementation and operation of a high quality teleradiology model increasing access to timely breast cancer assessment services for women in rural Australia. *BMC Health Serv Res* 2020; 20: 1103. DOI: 10.1186/s12913-020-05922-y
- [26] Sobez LM, Kim SH, Angstwurm M et al. Creating high-quality radiology reports in foreign languages through multilingual structured reporting. *Eur Radiol* 2019; 29: 6038–6048. DOI: 10.1007/s00330-019-06206-8
- [27] European Society of R. ESR paper on structured reporting in radiology. *Insights Imaging* 2018; 9: 1–7. DOI: 10.1007/s13244-017-0588-8
- [28] Kahn CE Jr., Heilbrun ME, Applegate KE. From guidelines to practice: how reporting templates promote the use of radiology practice guidelines. *J Am Coll Radiol* 2013; 10: 268–273. DOI: 10.1016/j.jacr.2012.09.025