Presentation Number: A576

Measuring Up: Systematic Review Of Measurement Tools For Post-Radiation Fibrosis In Patients With Head And Neck Cancer

Vanessa Helou, MD[2]; Purva H. Rumde, BA[1]; Pooja D. Reddy, BS[1]; Mary L. Klem, PhD, MLIS[2,3]; Matthew E. Spector, MD2[2]; Marci L. Nilsen, PhD, RN[2];; Kevin J. Contrera, MD MPH[2].

[1] School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213, USA; [2]Department of Otolaryngology-Head and Neck Surgery, University of Pittsburgh, Pittsburgh, PA 15213, USA.; [3] Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA 15213, USA.

BACKGROUND

- Long-term **radiation-induced fibrosis (RIF)** is a common and debilitating **late side effect** of head and neck cancer (HNC) treatment.[1]
- RIF can significantly impact patients' quality of life through **functional impairment** and **cosmetic changes**.
- Despite advancements in subjective and **objective measurement tools** to assess long-term fibrosis severity in these patients, there is a critical need for an **updated systematic review** to evaluate their **reliability, validity, and clinical utility**.

STUDY OBJECTIVE

To systematically review and evaluate the reliability, validity, and clinical utility of measurement tools for post-radiation fibrosis in patients with HNC.

METHODS

Design
- Systematic review following PRISMA guidelines.[2]

Databases Searched
- **Medline, Embase, Web of Science** and **Cochrane Library** (through November 2024).
- Restricted to English-language articles.

Study Selection
- **Population**: Patients with head and neck cancer after radiation therapy.
- **Interventions/Measures**: Imaging tools, physician-reported outcome measures, patient-reported outcomes measures (PROMs).
- **Outcomes**: Reliability, validity, and clinical utility of fibrosis measurement tools.
- **Excluded**: Abstracts, reviews, case reports, non-human participant studies.

Risk of Bias
- **COSMIN Reliability & Measurement Error checklist**: for clinician-reported outcomes (ClinRO) and Instrument-based tools.[3]
- **COSMIN RoB Checklist for PROMs v3.1**: for patient reported outcomes.[4]

RESULTS

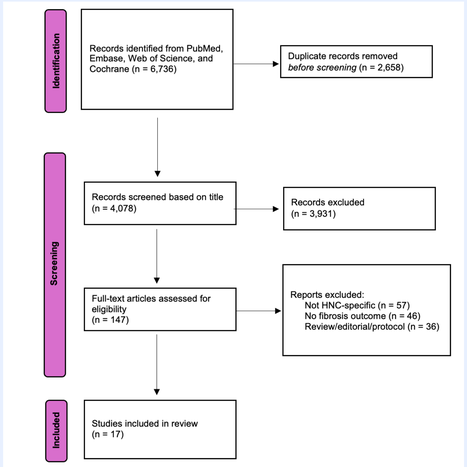

Figure 1: *PRISMA flow diagram for systematic reviews and meta-analyses.*

Study characteristics

- Out of 4078 articles, **17** included:
  - **10 imaging tools.**
  - **5 PROMs.**
  - **2 physician-reported outcome measures.**
- Median publication year: **2015**.
- Study design:
  - **Prospective designs:** 10 studies (59%).
  - **Cross-sectional:** 5 studies (29%).

Type of Measure	Measurement Tool	Publication	Clinical Evidence of Validity	Clinical Evidence of Reliability
PROMs	Neck Fibrosis Scale (NFS)	Noel 2021	Use of factor analysis during item reduction Face/content validity via expert/patient input	Test–retest ICC = 0.95 Cronbach's α = 0.95
		Saleem 2022	Performed factor analysis and confirmed a 2-factor structure (2 domains)	ICC >0.8 per domain Cronbach's α > 0.8 for each domain
	HN-LEF Symptom Inventory	Deng 2021	Content validity with patient and expert input	Cronbach' α >0.8 across domains
		Deng 2022	Construct validity confirmed with other validated measures (EORTC, QLQ-C30, NDI, SF-MPQ)	
	CTCAE adapted LENT-SOMA	Ho 2009	face/content validity only exploratory factor analysis for construct validity	Cronbach' α >0.85 overall
ClinROs	HN-ELAF	Deng 2015	Content validity from expert consensus and systematic item development	Inter- and intra-rater κ ≈ 0.7–0.8
		Deng 2016	Construct validity with clinican reporteed ratings compared with objective ultrasound tissue thickness	interrater κ = 0.70–0.95 for most fibrosis sites
Instrumental Tools	Young's Modulus (YM)	Zheng 2000	Compared YM of irradiated vs healthy patients	intra-/inter-observer reproducibility via %CV (±7–15%)
		Leung 2002	Construct validity as YM correlated with palpation grades, ROM, and symptoms	standardized protocol; prior repeatability cited
		Huang 2004	Descriminative validity as YM discriminated irradiated vs healthy tissue	
	Ultrasound elastography of sternocleidomastoid	Aulino 2018	Differentiated irradiated vs. non-irradiated tissue (higher stiffness in RT group (p < 0.05))	High inter- and intra-observer agreement (ICC > 0.9)
	MRI fibrosis volume (Vitrea™ contouring)	Che 2024	Fibrosis volume correlated with swallowing impairment on videofluoroscopy (Spearman ρ = 0.49–0.59, p < 0.01)	Inter-rater reliability achieved after consensus training (ICC = 0.95)
	MRI differentiation of post-RT fibrosis	Gong 1991	distinguish fibrosis vs. recurrence using T2 intensity patterns, though not perfectly specific	Serial imaging confirmation; reproducible signal intensity patterns; No formal κ/ICC reported
	Sonoelastography of salivary glands	Kaluzny 2014	ES values significantly higher in RT patients vs. controls; correlated with saliva volume and CTCAE xerostomia grade (p = 0.0005).	Reproducible elastography measures; strong correlation with functional outcomes, but no κ/ICC reported.
	IVIM diffusion/perfusion MRI	Lai 2013	Distinct diffusion and perfusion parameters for NPC vs. post-RT fibrosis (sensitivity/specificity up to 100%)	Quantitative parameters robust across subjects; reproducibility demonstrated
	Shear Wave Elastography (SWE) of SCM	Wen 2021	SWE correlated with LENT-SOMA fibrosis ((r = 0.557–0.649, p < 0.01)) grade and showed progression over 18 months	High inter- and intra-rater reliability (ICC > 0.75)
	Nakagami ultrasound imaging	Yang 2015	Differentiated asymptomatic vs. symptomatic fibrosis patients; correlated with RTOG scoring	Consistent quantitative imaging parameters across groups

Table 1: *Clinical evidence of validity and reliability for fibrosis measurement tools in head and neck cancer, as reported in the original studies. Evidence is summarized as presented by the authors, without application of COSMIN risk-of-bias criteria.*

Abbreviations*: NFS (Neck Fibrosis Scale), HN-LEF SI (Head and Neck Lymphedema and Fibrosis Symptom Inventory), CTCAE-adapted LENT-SOMA (Common Terminology Criteria for Adverse Events - Late Effects of Normal Tissues – Subjective, Objective, Management, Analytic scale),* HN-ELAF (Head and Neck External Lymphedema and Fibrosis assessment scale).

A.

Key. N/A = 0 Inadequate = 1 Doubtful = 2 Adequate = 3 Very Good = 4"	Type of Measure	Comparators	"Were patients stable in the time between the repeated fibrosis measurements?"	Was time between repeated fibrosis measurements appropriate?	Were the fibrosis measurement conditions similar for repeated measurements - except for the condition being a source of variation?	"Were the researchers adminstering fibrosis severly measurement w/o knowledge of repeated measurement scores in the same patient? "
Aulino et al. 2018	Instrument/Imaging	n/a	0	0	3	2
Che et al. 2024	Instrument/Imaging	swallowing function (VFS)	4	4	4	3
Deng et al. 2015	Clinician-Reported	n/a	3	0	3	2
Deng et al. 2016	Clinician-Reported	US grayscale	4	3	4	3
Gong et al. 1991	Instrument/Imaging	pathology (gold standard)	0	0	3	2
Huang et al. 2004	Instrument/Imaging	heathly vs irradiated tissue	0	0	3	2
Kaluzy et al. 2014	Instrument/Imaging	CTCAE xerostromia, saliva flow	0	0	3	2
Lai et al. 2013	Instrument/Imaging	histology (tumor vs fibrosis)	0	0	3	2
Leung et al. 2002	Instrument/Imaging	ClinRO palpation, ROM	2	0	4	2
Wen et al 2021	Instrument/Imaging	ClinRO LENT-SOMA	4	4	4	2
Yang et al. 2015	Instrument/Imaging	ClinRO RTOG fibrosis	0	0	3	2
Zheng et al. 2000	Instrument/Imaging	ClinRO palpation, ROM	4	4	4	2

B.

Key. N/A = 0 Inadequate = 1 Doubtful = 2 Adequate = 3 Very Good = 4"	Type of Measure	Comparators	"Were the researchers adminstering fibrosis severly scores w/o knowledge of repeated measurement scores in the same patient? "	Did the study have other important design/statistical method flaws?	"For ordinal scores: was weighted kappa calculated? "	For continuous scores: was intraclass correlation coefficient (ICC) calculated?	For dichotamous/nominal scores: was kappa calculated for each category against the other catgeories combined?
Aulino et al. 2018	Instrument/Imaging	n/a	2	1	1	1	0
Che et al. 2024	Instrument/Imaging	swallowing function (VFS)	4	3	0	4	0
Deng et al. 2015	Clinician-Reported	n/a	2	2	2	0	1
Deng et al. 2016	Clinician-Reported	US grayscale	3	3	2	0	1
Gong et al. 1991	Instrument/Imaging	pathology (gold standard)	2	2	1	1	1
Huang et al. 2004	Instrument/Imaging	heathly vs irradiated tissue	2	2	1	1	0
Kaluzy et al. 2014	Instrument/Imaging	CTCAE xerostromia, saliva flow	2	2	1	1	0
Lai et al. 2013	Instrument/Imaging	histology (tumor vs fibrosis)	2	2	0	1	0
Leung et al. 2002	Instrument/Imaging	ClinRO palpation, ROM	2	2	1	1	0
Wen et al 2021	Instrument/Imaging	ClinRO LENT-SOMA	4	3	1	4	0
Yang et al. 2015	Instrument/Imaging	ClinRO RTOG fibrosis	3	2	1	1	0
Zheng et al. 2000	Instrument/Imaging	ClinRO palpation, ROM	4	2	0	1	0

C.

NA: Not applicable -: Indicates "Doubtful." ?: Indicates "Inadequate." ±: Indicates "Adequate." +: Indicates "Very Good."	PROM (Index Measure)	Prom development 1. design 2. comprehensibility 3. comprehensiveness	Content Validity 1. Relevance 2. Comprehensiveness 3. Comprehensibility	Structural Validity 1. statistical methods	Internal Consistency	Hypothesis testing	Reliability	Measurement Error	Criterion Validity	Overall Rating
Deng et al. 2021	HN-LEF Symptom Inventory (refinement)	+	+	±	+	?	?	?	NA	?
Deng et al. 2022	HN-LEF Symptom Inventory (construct validity)	NA	NA	NA	?	+	?	?	NA	?
Ho et al. 2009	Patient late toxicity PROM (adapted LENT-SOMA→ CTCAE)	±	±	±	±	?	?	?	NA	?
Noel et al. 2021	Neck Fibrosis Scale (PROM)	+	±	±	+	?	+	?	NA	?
Saleem et al. 2022	Neck Fibrosis Scale (PROM, domain validation)	NA	NA	±	+	?	+	?	NA	?

Table 2: *COSMIN risk of bias ratings for measurement tools assessing post-radiation fibrosis in head and neck cancer. **A**. and **B.** summarize clinician-reported outcome measures (ClinROs), and instrumental tools. **C** summarizes ratings for patient-reported outcome measures (PROMs) Ratings were assigned using COSMIN risk of bias checklists (v3.1 for PROMs; reliability and measurement error checklist for ClinROs/Instruments), applying the "worst score counts" method per measurement property for **C.** These risk of bias ratings are distinct from the clinical evidence of validity and reliability reported by the original studies (see **Table 1**).*

CONCLUSIONS

- Multiple **subjective** RIF measurement tools **show evidence of clinical reliability and validity.** However, per COSMIN's guidelines, they are rated as **'inadequate' due to incomplete reporting** of **measurement error statistics**.
- Among **objective** RIF measurement tools, **Young's modulus showed strong clinical validity and reliability,** but did not meet COSMIN criteria for reliability due to coefficient of variation (%CV) reported over intraclass correlation coefficient.
- Further development of **low-bias, RIF measurement tools meeting COSMIN criteria** is critically needed.

REFERENCES

1. Brook I. Late side effects of radiation treatment for head and neck cancer. Radiat Oncol J 2020;38(2):84-92. (In eng). DOI: 10.3857/roj.2020.00213.
2. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj 2021;372:n71. (In eng). DOI: 10.1136/bmj.n71.
3. LB Mokkink, M Boers, CPM van der Vleuten, LM Bouter, J Alonso, DL Patrick, HCW deVet, CB Terwee. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Medical Research Methodology. 2020;20(293)
4. LB Mokkink, E Elsman, CB Terwee. (2024). The COSMIN guideline systematic reviews of Patient-Reported Outcome Measures (PROMs). Qual Life Res. https://doi.org/10.1007/s11136-024-03761-6.