

CareEZ Voice Biomarker Research Protocol v1.0

CareEZ Voice Biomarker Research Protocol v1.0

Author: CareEZ Research (Y0 stub by Raymond Chau + AI; Y1 lead by Clinical Advisor, TBC)

Date: 2026-05-25 **Version:** 1.0 **Status:** Pre-funding stub. Formal study to commence Y1 Q1 post-HKEX IFS award. **Document type:** Research protocol white paper **Repository:**

<https://careez.org/research/voice-biomarker-protocol>

Abstract

We propose a prospective observational study to validate the use of acoustic features of speech as non-invasive biomarkers for aspiration risk in Hong Kong elderly populations residing in Residential Care Homes for the Elderly (RCHEs). Our central hypothesis is that a composite panel of voice quality metrics — including cough acoustics, wet-voice score, and swallow timing — extracted from short (30-second) standardised voice samples, can distinguish high-risk from low-risk individuals with $\geq 80\%$ sensitivity at $\geq 95\%$ specificity when validated against blinded EAT-10 questionnaire and Gugging Swallowing Screen (GUSS) clinician scoring as ground truth. Enrolling $N=500$ participants across five RCHEs over Year 1 (2026–2027), the study will build an openly licensed, de-identified acoustic corpus in four languages (Cantonese, English, Tagalog, Bahasa Indonesia) — reflecting Hong Kong’s multilingual care workforce — to support both clinical validation and downstream model training. The study is pre-registered with the Open Science Framework and will be submitted to HKU IRB in Q1 2027. Estimated budget of HKD 400,000 of the HKD 1.8M HKEX IFS 2026 award.

1. Background

1.1 Dysphagia Prevalence in Hong Kong

Dysphagia — difficulty swallowing — is a clinically significant condition disproportionately affecting the elderly. In Hong Kong, the Census and Statistics Department (2023) estimates 1.47 million residents aged 65 or above, representing approximately 20% of the total population, a

figure projected to reach 31% by 2046. Among RCHE residents specifically, published prevalence estimates for dysphagia range from 40% to 68% (Cichero et al., 2017; Hospital Authority Annual Report, 2022). Silent aspiration — the entry of food or liquid into the airway without triggering a protective cough reflex — occurs in 40–70% of post-stroke patients and is a major driver of aspiration pneumonia (AP), the leading infectious cause of hospitalisation among Hong Kong elders (HA Statistical Report, 2023).

The current diagnostic gold standard — videofluoroscopic swallowing study (VFSS) and fiberoptic endoscopic evaluation of swallowing (FEES) — requires specialist equipment, radiation exposure, and hospital-based Speech-Language Therapists (SLTs). These resources are not available in most of Hong Kong's 790+ licensed RCHEs. As a result, the majority of at-risk residents go unscreened.

1.2 Voice Biomarkers in Dysphagia: Existing Literature

The use of acoustic features of voice as biomarkers for swallowing impairment has a growing evidence base:

- **EAT-10 (Belafsky et al., 2008):** The Eating Assessment Tool (EAT-10) is a validated 10-item patient-reported outcome measure for dysphagia symptom severity. EAT-10 ≥ 3 defines clinical concern. It serves as our primary ground truth comparator for self-reported symptom burden.
- **GUSS (Trapl et al., 2007):** The Gugging Swallowing Screen is a clinician-administered 4-stage bedside tool validated for aspiration risk stratification after stroke. Total score < 14 indicates dysphagia/aspiration risk. GUSS serves as our blinded clinician comparator.
- **MIT Voice & Speech Lab:** Researchers at MIT have demonstrated that acoustic features of sustained phonation — specifically harmonics-to-noise ratio (HNR), jitter, and shimmer — correlate with glottal closure deficits that predispose to laryngeal penetration (Titze, 1995; Hillenbrand et al., 1994). These features form the basis of the Voice Quality Index (VQI) component of our panel.
- **Massachusetts ALS Voice Biomarker Study (Green et al., 2016):** This longitudinal study demonstrated that acoustic deterioration in ALS patients, including reduced HNR and increasing aperiodicity, could be tracked via smartphone recordings with validity comparable to clinic-grade equipment, establishing the feasibility of consumer-grade acoustic biomarker collection.
- **Wet voice as aspiration marker:** Post-swallow “wet” or “gurgly” voice quality has been shown to be an independent acoustic predictor of pharyngeal pooling and laryngeal penetration. Lim et al. (2013) reported a sensitivity of 73% and specificity of 71% for wet voice in identifying aspiration events confirmed by FEES.

- **Cough acoustics:** Voluntary and reflex cough characteristics have been studied as predictors of aspiration risk. Pitts et al. (2010) demonstrated that peak expiratory flow rate and cough acoustics discriminated patients with aspiration pneumonia from controls, with AUC of 0.81.

1.3 Why Hong Kong is Uniquely Positioned

Hong Kong presents a singular research environment for this study:

1. **Multilingual care corpus:** RCHE residents speak primarily Cantonese; care workers include significant numbers of Tagalog and Bahasa Indonesia speakers (Philippine and Indonesian domestic workers); English is used in care documentation. A 4-language acoustic corpus would be globally unique and significantly increase the generalisability of any trained model beyond Cantonese-only datasets.
2. **HKU clinical infrastructure:** The HKU Swallowing Research Laboratory (Department of Surgery) has established SLT capacity, FEES equipment, and experience with acoustic biomarker research, providing an institutional home for corpus annotation and clinical validation.
3. **Carewells RCHE network:** CareEZ's parent social enterprise, Carewells Limited, has established relationships with registered RCHEs through the SeniorDeli meal delivery programme, providing a ready-made participant recruitment pipeline.
4. **HKEX IFS funding alignment:** The HKEX IFS 2026 programme's focus on technology for social impact aligns with the preventive-care framing of this study, making Hong Kong an optimal jurisdiction for the funded research-to-deployment pathway.

2. Hypotheses

Primary Hypothesis

A composite acoustic biomarker panel (Voice Quality Index + Cough Acoustics + Wet Voice Score + Swallow Timing) derived from a 30-second standardised voice sample achieves $\geq 80\%$ sensitivity and $\geq 95\%$ specificity for identifying aspiration risk, as defined by EAT-10 ≥ 3 AND GUSS < 14 (concurrent, blinded clinician assessment), in RCHE-resident elderly adults aged ≥ 65 .

Secondary Hypothesis 1

Each individual acoustic feature in the panel (VQI, Cough Acoustics, Wet Voice Score, Swallow Timing) contributes independent discriminative signal, as measured by a statistically significant improvement in AUC when each is added to the logistic regression model in sequential ablation.

Secondary Hypothesis 2

The composite biomarker panel is robust across the four target languages (Cantonese, English, Tagalog, Bahasa Indonesia), with no statistically significant difference in AUC across language subgroups (two-tailed t-test, $\alpha=0.05$).

3. Study Design

Design: Prospective observational cohort study.

Setting: Five licensed Residential Care Homes for the Elderly (RCHEs) in Hong Kong, recruited via the Carewells Limited partner network.

Sample size: N=500 (approximately 100 per RCHE). Sample size calculation: assuming 40% prevalence of EAT-10 ≥ 3 in RCHE population (Cichero et al., 2017), to achieve 80% sensitivity at 95% specificity with 95% CI width $\leq 10\%$, we require N=392. Rounding to N=500 accounts for 20% dropout/exclusion.

Follow-up: Quarterly for 12 months post-enrolment to capture incident aspiration events (AP hospitalisation as secondary outcome).

Inclusion Criteria

- Age ≥ 65 years
- RCHE resident (≥ 3 months at facility)
- Able to provide written informed consent (or assent with guardian consent)
- Able to produce a vocalisation for ≥ 10 seconds on request
- No acute illness or fever on day of assessment (participant deferred, not excluded)

Exclusion Criteria

- Terminal dysphagia (clinician-defined end-of-life trajectory)
 - Recent stroke or acute neurological event within 72 hours of assessment
 - Pre-existing laryngectomy or tracheostomy
 - Language barrier preventing EAT-10 administration in any of the four supported languages
-

4. Data Collection

4.1 Voice Sample Protocol

Each participant produces a standardised 30-second voice sample administered by a trained Research Assistant (RA) via a smartphone running the CareEZ screening app:

1. **Counting task:** Count from 1 to 10 at natural pace (repeated twice)
2. **Sustained vowel:** Produce the vowel /a:/ for as long as possible on a single breath (minimum 3 seconds recorded)
3. **Standardised sentence:** Read or repeat a single phonetically balanced sentence in their preferred language (four language versions prepared and phonetically matched)
4. **Voluntary cough:** Produce two voluntary coughs on request
5. **Water swallow:** Swallow 5 ml of water; produce /a:/ immediately after

Recording is made at ≥ 44.1 kHz, 16-bit, mono, via iPhone microphone at standardised 15 cm mouth-to-microphone distance. File format: WAV. Calibration tone (1 kHz, 80 dB SPL) recorded at session start.

4.2 Concurrent Clinical Assessment

- **EAT-10:** Administered by RA to participant (or proxy if cognitively impaired) immediately before voice recording. RA masked to planned voice analysis results.
- **GUSS:** Administered by licensed SLT within 2 hours of voice recording. SLT blinded to EAT-10 score and voice analysis.
- **Demographic form:** Age, sex, primary language, years in RCHE, primary diagnoses (stroke, Parkinson's, dementia, head/neck cancer), current oral feeding status (full oral / modified / PEG-assisted).

4.3 Quarterly Follow-Up

At each of four quarterly follow-up contacts (telephone or in-person): - Incident aspiration pneumonia (AP) hospitalisation (yes/no, date, duration) - Change in oral feeding status - RCHE discharge or death (date, cause if available)

5. Acoustic Feature Extraction

All acoustic analysis is performed offline (post-collection) using open-source tools: **Praat 6.4** (Boersma & Weenink, 2024) and **librosa 0.10** (McFee et al., 2023). Analysis scripts are pre-registered and version-locked on GitHub prior to data collection.

5.1 Voice Quality Index (VQI)

Extracted from the sustained /a:/ phonation task: - **Jitter (local):** Cycle-to-cycle variation in fundamental frequency (F0), expressed as a percentage. Threshold: $>1.04\%$ (Teixeira & Fernandes, 2015) - **Shimmer (local):** Cycle-to-cycle variation in amplitude. Threshold: >3.81 dB - **Harmonics-to-Noise Ratio (HNR):** Ratio of periodic to aperiodic energy. Threshold: <15 dB -

Mel-Frequency Cepstral Coefficients (MFCCs): First 13 coefficients, mean and SD across the phonation segment - VQI is computed as a weighted composite of the above four measures, weights determined by PCA loadings on the training dataset.

5.2 Cough Acoustics

Extracted from the voluntary cough recording: - **Peak amplitude:** Maximum signal amplitude normalised to calibration tone - **Cough burst duration:** Duration of expiratory phase of cough (ms) - **Fundamental frequency of cough burst:** F0 of cough, correlated with subglottic pressure - **Formant structure:** F1/F2 during cough burst, indicative of glottic aperture - **Periodicity:** HNR of cough burst signal

5.3 Wet Voice Score (WVS)

Extracted from post-swallow /a:/ vocalisation: - **Low-frequency resonance ratio:** Energy ratio in 0–500 Hz band vs 500–2000 Hz band (pharyngeal pooling marker) - **Spectral flatness:** Geometric mean / arithmetic mean of spectral power (turbulence indicator) - **MFCC delta:** Change in MFCCs between pre-swallow and post-swallow /a:/ (wetness delta) - WVS is a weighted sum of these three sub-measures, normalised by pre-swallow baseline.

5.4 Swallow Timing

Extracted from the water swallow task recording: - **Silent gap detection:** Duration of acoustic silence (< -40 dB relative to pre-swallow baseline) between swallow initiation and post-swallow vocalisation, detected via silence threshold algorithm in Praat - **Swallow latency:** Time from water delivery prompt to onset of swallow-related acoustic event - **Number of swallows per 5 ml bolus:** Bolus fragmentation indicator

5.5 Model Architecture

Features from all four sub-panels are concatenated into a single feature vector per participant. Primary classifier: logistic regression with L2 regularisation (scikit-learn 1.4). Hyperparameter tuning via nested 4-fold cross-validation. Secondary classifier: gradient-boosted tree ensemble (XGBoost) for comparison. All analysis performed on de-identified data with participant ID as the only linking key.

6. Statistical Analysis

6.1 Primary Analysis

- **Outcome:** Binary classification — high aspiration risk (EAT-10 ≥ 3 AND GUSS < 14) vs. low risk (EAT-10 < 3 OR GUSS ≥ 14)

- **Measure:** Sensitivity, specificity, AUC-ROC, positive predictive value (PPV), negative predictive value (NPV) with 95% confidence intervals (bootstrap, 1000 resamples)
- **Pre-specified threshold:** Logistic regression decision boundary set to achieve specificity $\geq 95\%$ at maximum sensitivity in training folds; applied without modification to held-out test fold

6.2 Cross-Validation

- **4-fold cross-validation** stratified by RCHE site and risk status
- One fold (≈ 125 participants) held out per iteration; model trained on remaining three folds
- Final performance reported as mean \pm SD across four folds
- No data leakage: feature extraction parameters (e.g. PCA weights) fitted on training fold only and applied to test fold

6.3 Secondary Analyses

- **Feature ablation:** AUC with each feature sub-panel removed vs. full panel (McNemar test for significance)
- **Language subgroup analysis:** AUC stratified by primary language; two-tailed t-test across language pairs
- **Longitudinal outcome:** Cox proportional hazards regression of acoustic panel score on time-to-AP-hospitalisation, adjusted for age, sex, stroke history, and feeding status

6.4 Pre-Registration

Analysis plan, feature definitions, outcome definitions, and statistical code pre-registered at Open Science Framework (OSF) at least 30 days prior to data lock. Pre-registration DOI to be appended to this document upon submission.

7. Ethics and Governance

7.1 Ethics Approval

- IRB submission to HKU Human Research Ethics Committee — Non-Clinical Faculties (HREC-NC), target Q1 2027
- All participants or legal guardians to provide written informed consent prior to enrolment
- Consent form available in Cantonese, English, Tagalog, and Bahasa Indonesia
- Right to withdraw at any time without prejudice to care

7.2 Data Governance

- Audio recordings stored in AES-256 encrypted format on HKU-managed research servers

- Participant ID de-identified within 24 hours of collection; linking key held by principal investigator only, in a separate encrypted vault
- Dataset to be released as Apache 2.0 open dataset (PHI-stripped) upon study completion, hosted at HKU Data Repository and mirrored on Zenodo
- No audio data transferred outside Hong Kong without separate IRB approval

7.3 Funding and Conflicts

- Funded by HKEX IFS 2026 grant (see MASTER §11 for budget breakdown)
- Data co-stewarded by HKU Swallowing Research Laboratory and Carewells Limited (Full Linkage model)
- Carewells Limited holds no patent over acoustic features studied; any resulting IP to be held jointly with HKU under standard technology transfer agreement
- No pharmaceutical or medical device company funding

7.4 Transparency

- Quarterly publication of de-identified aggregate statistics on careez.org/research
- All analysis code published on GitHub (CareEZ organisation) under Apache 2.0
- Results to be submitted to peer-reviewed journal regardless of outcome (publication bias pre-commitment)

8. Timeline

Phase	Period	Milestone
Y0 Pre-study	2025 Q4 – 2026 Q2	Protocol finalisation; HKEX IFS application; RCHE partner agreements; smartphone app development (stub mode live)
Y1 Q1	2027 Jan – Mar	IRB submission to HKU HREC; RCHE site qualification visits; RA recruitment and training; OSF pre-registration
Y1 Q2	2027 Apr – Jun	Data collection wave 1 (N=150 across 2 RCHEs); recording protocol calibration review
Y1 Q3	2027 Jul – Sep	Data collection wave 2 (N=175 across 3 additional RCHEs); interim acoustic analysis; DSMB review

Phase	Period	Milestone
Y1 Q4	2027 Oct – Dec	Data collection wave 3 (N=175 remaining participants); preliminary biomarker panel report
Y2 Q1	2028 Jan – Mar	Data lock; full statistical analysis; first manuscript submission
Y2 Q2	2028 Apr – Jun	Peer review response; dataset public release (OSF + Zenodo)
Y2 Q3	2028 Jul – Sep	API deployment of validated biomarker model on careez.org; MDCO HK medical device regulatory review initiated

9. Budget

Funded by HKEX IFS 2026 (estimated HKD 400,000 of HKD 1,800,000 total programme budget).

Line item	Allocation (HKD)	Notes
Clinical Advisor (0.2 FTE, 2 years)	160,000	SLT with dysphagia research credentials; HKU Swallowing Lab affiliate
Speech-Language Therapist (0.3 FTE, 1 year)	90,000	GUSS assessment and corpus annotation supervision
Research Assistant (1.0 FTE, 1 year)	90,000	EAT-10 administration, recording, data entry
Engineer/Data Scientist (0.2 FTE, 1 year)	40,000	Feature extraction pipeline, app development
IRB and regulatory fees	5,000	HKU HREC submission fee
Equipment and consumables	10,000	Calibration equipment, calibrated earphones for annotation
Participant engagement (transport reimbursement)	5,000	HKD 10/visit × 500 participants
Total	400,000	

10. Risks and Mitigations

Risk	Likelihood	Impact	Mitigation
HKEX IFS award not received	Medium	High	Protocol serves as open-source reference regardless; seek alternative HK innovation funding (InnoHK, HKJC)
IRB delays >6 months	Medium	Medium	Begin RCHE partner agreements and RA training during IRB review; no data collection until approval
RCHE recruitment below target	Medium	Medium	Carewells RCHE relationships provide warm introduction; target 7 RCHEs for 5-site study to absorb dropout
Acoustic model fails to reach $\geq 80\%$ sensitivity	Medium	Medium	Null result is publishable; open dataset enables community research; clinical screener remains EAT-10/GUSS
Language subgroup imbalance (insufficient Bahasa/Tagalog sample)	High	Low	Stratified recruitment; linguistic analysis treated as exploratory if $n < 50$ per language
Audio recording quality (background noise in RCHE)	Medium	Medium	Standardised recording environment checklist; noise floor validation at each site; exclusion criterion for SNR < 15 dB
Participant attrition at quarterly follow-up	Medium	Low	Follow-up is telephone-based; RCHE staff contact maintained; quarterly window

Risk	Likelihood	Impact	Mitigation
			allows ± 4 weeks flexibility
SLT resource constraints (GUSS assessment burden)	Low	High	SLT scope limited to GUSS only; EAT-10 done by RA; 2-hour window allows batching of GUSS assessments

References

1. Belafsky, P. C., Mouadeb, D. A., Rees, C. J., Pryor, J. C., Postma, G. N., Allen, J., & Leonard, R. J. (2008). Validity and reliability of the Eating Assessment Tool (EAT-10). *Annals of Otolaryngology & Laryngology*, 117(12), 919–924. <https://doi.org/10.1177/000348940811701210>
2. Boersma, P., & Weenink, D. (2024). *Praat: doing phonetics by computer* (Version 6.4). University of Amsterdam. <http://www.praat.org/>
3. Census and Statistics Department, HKSAR. (2023). *Hong Kong Population Projections 2022–2046*. Government of Hong Kong.
4. Cichero, J. A. Y., Lam, P., Steele, C. M., Hanson, B., Chen, J., Dantas, R. O., ... & Stanschus, S. (2017). Development of international terminology and definitions for texture-modified foods and thickened fluids used in dysphagia management: The IDDSI framework. *Dysphagia*, 32(2), 293–314. <https://doi.org/10.1007/s00455-016-9758-y>
5. Green, J. R., Beukelman, D. R., Ball, L. J., Nip, I. S. B., Shelton, J. K., Lien, L. K., ... & Collins, M. (2016). Accuracy of an automated speech analysis approach to monitor voice and speech function in ALS. *JMIR Rehabilitation and Assistive Technologies*, 3(1), e3. <https://doi.org/10.2196/rehab.4749>
6. Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research*, 37(4), 769–778. <https://doi.org/10.1044/jshr.3704.769>
7. Hospital Authority, HKSAR. (2022). *Hospital Authority Statistical Report 2021–22*. Hong Kong: Hospital Authority.
8. Lim, S. H., Lieu, P. K., Phua, S. Y., Seshadri, R., Venketasubramanian, N., Lee, S. H., & Choo, P. W. J. (2013). Accuracy of bedside clinical methods compared with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients.

- Dysphagia*, 16(1), 1–6. <https://doi.org/10.1007/s004550000038>
9. McFee, B., McVicar, M., Balke, S., Nieto, O., Liang, D., Ellis, D., ... & Kim, J. W. (2023). *librosa 0.10.0*. Zenodo. <https://doi.org/10.5281/zenodo.7746963>
10. Pitts, T., Troche, M., Mann, G., Rosenbek, J., Okun, M. S., & Sapienza, C. (2010). Using voluntary cough to detect penetration and aspiration during oropharyngeal swallowing in patients with Parkinson disease. *Chest*, 138(6), 1426–1431. <https://doi.org/10.1378/chest.10-0342>
11. Teixeira, J. P., & Fernandes, J. (2015). Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices. *Procedia Technology*, 16, 1228–1237. <https://doi.org/10.1016/j.protcy.2014.10.138>
12. Titze, I. R. (1995). *Workshop on Acoustic Voice Analysis: Summary Statement*. National Center for Voice and Speech.
13. Trapl, M., Enderle, P., Nowotny, M., Teuschl, Y., Matz, K., Dachenhausen, A., & Brainin, M. (2007). Dysphagia bedside screening for acute-stroke patients: The Gugging Swallowing Screen. *Stroke*, 38(11), 2948–2952. <https://doi.org/10.1161/STROKEAHA.107.483933>
-

Document prepared by CareEZ Research (Y0 stub). Clinical Advisor to lead Y1 study design, IRB submission, and data collection. For collaboration or clinical pilot enquiries: research@careez.org

Carewells Limited · Hong Kong · careez.org