Ambient Voice Technology (AVT) Measurement Toolkit v1.0

A collaboration between NHS England and Health Innovation South West

A benefits realisation management approach

Digital tools rarely deliver benefits just by being deployed. Benefits can depend on:

- Workflow change
- Behaviour change
- Adoption and sustained use
- Local context

Ambient Voice Technology (AVT) benefits (e.g. time savings, wellbeing, data quality, productivity gains) only materialise if:

- Templates are well designed
- Clinicians trust the outputs
- Editing burden does not outweigh gains
- Deployment replaces other admin (e.g. outsourced typing)

A benefits realisation management approach treats benefits as managed assets, not automatic consequences of delivery:

- **Design the conditions under which benefits can emerge**
- **Evaluate whether those conditions are being met**

The logic of identifying profiling, baselining, tracking and sustaining across clinical, operation, workforce and patient domains is aligned to the Government Project Delivery Teal Book[1].

1. About The Teal Book - Government Project Delivery

Background

The South West NHSE team has worked with the Health Innovation Network SW to develop this Measurement Toolkit to support Ambient Voice Technology (AVT) deployment. The aim has been to provide a framework and associated tools to specifically:

- define and categorise benefits in a consistent manner in alignment with The Teal Book[1];
- provide a set of tools that can be further tested in practice to measure and evaluate AVT benefits.

What it is

AVT evaluation is the systematic assessment of whether ambient voice technology delivers real, safe, and sustainable benefit in clinical practice, by testing its impact on quality, workload, risk, and experience under real-world NHS conditions.

Therefore, measuring the impact of AVT is about checking, in real NHS settings, whether these tools genuinely help staff and patients without causing new problems or risks.

Why it matters

The NHS can make defensible, scalable decisions about AVT that deliver real benefits to patients and staff, without introducing unmanaged risk or false confidence.

In this document, aspirational target areas have been provided that reflect research studies and AVT supplier information. ***Please note: these are not NHSE targets.***

1. About The Teal Book - Government Project Delivery

Benefits Expectations, Assumptions and Dependencies

- The benefits outlined in this Toolkit (v1.0) are primarily aligned to improvements in frontline productivity, reflecting a system-level expectation that digital transformation should support more efficient and effective use of clinical and administrative time.

- We have developed it to avoid duplication of effort and to enable local discretion over ***what*** benefits are measured and ***how*** they are evidenced, recognising variation in service context, maturity, priorities, and available data.

- Users are therefore encouraged to select benefit measures that are meaningful, proportionate, and feasible within their local setting, while maintaining alignment with the overarching productivity intent.

- Going forward, recipients of Frontline Productivity (FP) investment funds will be required as part of the Memorandum of Understanding (MoU) to select and measure defined benefits (still to be determined).

- **Note to users:** We would like this toolkit to be iteratively developed and improved based on your feedback. We are keen to work with providers to gain insights into the best way to collect data, whether the tools provide valuable feedback and if there are other better tools that could be added. Please email: england.swaicollaboration@nhs.net if you would like to be directly involved in this process.

How we defined the Four Domains

Identification and defining the four domains was completed in a clear, practical sequence reflecting good practice in Benefits realisation (Managing Benefits™) and NHS Digital transformation:

Clarify the problem & purpose - Sustainable workforce wellbeing and Productivity: aspirational targets mentioned by suppliers, and research and early evaluation studies

What isn't working today, for whom, and why change is needed



Align to strategy & policy - Fit for the Future

Linked to organisational, system, and national priorities



Review evidence & literature

Used research, evaluations, and comparable case studies to identify plausible benefits



Engage stakeholders - the early adopters

Surfaced context-specific benefits from service delivery and their patient perspectives



Structure benefits clearly

Grouped benefits into meaningful categories (e.g. patient, staff, productivity, safety, system)



Prioritised what matters most - reduced into the areas of quality, experience and productivity

Focused on benefits that are material, achievable, and evidence-informed, and aligned to Fit for the Future strategy



Define success at a high level - used the aspirational measurements made by suppliers and recent studies such as the GOSH study

Described what success would look like and how it could be measured in the Toolkit

Ambient Voice Technology (AVT) Measurement Toolkit

	A. DATA QUALITY	B. PATIENT EXPERIENCE	C. WORKFORCE EXPERIENCE	D. PRODUCTIVITY
DOMAINS	Hypothesis: AVT maintains or improves accuracy/completeness without raising incidents.	Hypothesis: AVT improves consultation experience without harming privacy perceptions.	Hypothesis: AVT lowers cognitive burden and notetaking stress, improving job satisfaction.	Hypothesis: Time saved increases capacity without extending hours.
SUB-DOMAINS	Documentation quality, completeness & safety Equality, inclusion & accessibility	Patient experience and acceptability	Workforce experience and acceptability Workforce wellbeing	Clinician efficiency & documentation time Clinic efficiency (throughput) Financial impact & value for money



A. Data Quality

SUB-DOMAINS

Documentation quality, completeness & safety

Hypothesis: AVT maintains or improves the accuracy and completeness of documentation for all patients.

Equality, inclusion & accessibility

Hypothesis: AVT does not disadvantage diverse users and supports inclusive care.

Aspirational target

- Error reduction in clinical notes (data quality)
- Standardisation of workflows
- Sustained productivity gains

MEASUREMENT

Template adherence

Measure: Degree to which clinicians complete required sections of standard documentation templates.
Tool: ***Automated tools are currently not available for this.***

Documentation accuracy by accent/language

Measure: Transcription error rate by accent/language - review a sample of transcripts for patients with different accents or languages and calculate the proportion of notes containing clinically meaningful transcription inaccuracies within each accent/language group.
Tool: Sheffield Assessment Instrument for Letters (SAIL) Referral Letter Assessment Tool (SAIL)

Rework rate

Measure: The amount of notes that require corrections, edits, or rewrites, and the significance of the editing.
Tool: **Single Likert scale question**, for example: 'How much editing did this note require prior to sign-off?' ('no editing' → 'extensive editing'). Proportions for each response category could then be calculated.
Tool: **Word Error Rate (WER)**
Tool (for a broader measure of document quality): **PDQI-9** Physician Documentation Quality Instrument (PDQI-9) tool

B. Patient experience

SUB-DOMAINS

Patient experience and acceptability

Hypothesis: AVT use is acceptable to the patient and improves consultation experience for both patients and staff members.

MEASUREMENT

Patient experience

Measure: Patient-reported experience of receiving healthcare utilising AVT (e.g. communication with clinician, feeling listened to, quality of interactions).
Tool: AVT-specific patient survey or qualitative interviews. Consider pre-existing local data collection of patient experience data.
Tool: **The Consultation and Relational Empathy (CARE) Measure** – a 10-item questionnaire to assess relational quality. The CARE Measure Website / **The Net Promoter Score (NPS)**

Patient acceptability for AVT

Measure: How patients feel about the use of AVT (e.g. comfort, trust, and perceived usefulness) and their agreement for its use to capture and process their spoken interactions.
Tool: **% patients consenting to AVT use** - percentage of patients who agree to the use of AVT after being informed.
Tool: AVT-specific patient survey to capture patient attitudes towards AVT, or patient interviews.

Aspirational target

- Better quality interaction
- Improved patient experience
- Improved adherence and condition management related to clearer communications and documentation

C. Workforce experience

SUB-DOMAINS

MEASUREMENT

Workforce experience and acceptability

Hypothesis: AVT lowers cognitive burden and notetaking stress, improving job satisfaction.

Activation rate

Measure: Staff-reported acceptability and experience of utilising AVT as part of clinical practice (e.g. trust and comfort in the technology, fit with workflow).
Tool: AVT-specific staff survey or qualitative interviews.
Tool: Clinical system activity reporting / System Usability Scale Calculator

Uncontracted time worked

Measure: Total time clinicians spend working outside of their contracted hours.
Tool: **Number of minutes worked outside of contracted hours** - additional minutes worked outside of usual working hours.
Tool: Self-reported data collection.

Workforce wellbeing

Hypothesis: The benefits of AVT reduce stress, burnout, and absenteeism, while improving work-life balance and job satisfaction.

Stress and burnout

Measure: Clinician burnout and workplace stress.
Tool: **Mini-Z** – a 10-item measure assessing burden, stress and satisfaction and drivers of burnout / Maslach Burnout Inventory Test / Oldenburg Burnout Inventory / Copenhagen Burnout Inventory

Job satisfaction

Measure: Clinician's overall degree of contentment with their job.
Tool: **Mini-Z** Q1 of this 10-item measure assesses job satisfaction, 'Overall, I am satisfied with my current job' (agree strongly → strongly disagree).Mini-Z Survey

Cognitive load

Measure: Mental effort required to perform a task.
Tool: Subjective workload rating that measures mental, physical, and temporal demand, and performance, effort, and frustration level during tasks. The measure should be anchored to and administered directly after the individual has performed a selected task.
Tool: NASA Task Load Index (NASA-TLX) / System Usability Scale Calculator

Aspirational target

- Reduced documentation burden
- Reduced after hours documentation
- Improved job satisfaction & wellbeing
- Improved staff retention

D. Productivity

SUB-DOMAIN

MEASUREMENT

Clinical efficiency & documentation time

Hypothesis: AVT reduces the time to create, edit and finalise notes.

Documentation time – clinical notes

Measure: The time taken by a clinician to create clinical notes.
Tool: **Minutes per note** – the number of minutes a clinician spends creating and finalising a single clinical note, grouped by note type where applicable.
Tool: Self-reported

Total documentation time per session

Measure: The total time a clinician spends on documentation during a session (e.g. clinic or shift). This includes clinical notes and other paperwork such as referral letters and reports.
Tool: **Total documentation minutes per session** – total number of minutes spent on documentation during a clinic or shift.
Tool: Clinical system activity reporting / **TimeCaT**

Notes closed same day

Measure: The proportion of clinical notes completed and signed off on the same day as the consultation.
Tool: **% notes closed same day** - in an Electronic Patient Record (EPR), this can be measured by comparing each note's consultation date/time with its 'note signed' timestamp and counting how many were signed on the same calendar day.
Tool: The proportion can be calculated as: (number signed same day ÷ total day's notes) × 100.

Clinic efficiency (throughput)

Hypothesis: Time saved by using AVT improves efficiency, which may allow for increased capacity without extending hours.

Clinic overrun minutes

Measure: The amount of time a clinic runs beyond its scheduled end.
Tool: **Minutes of clinic overrun** – the difference between the scheduled clinic end time and the time the final patient is seen or the last note is completed.
Tool: If the final patient seen or last note completed is beyond the scheduled clinic end time, the difference represents the number of minutes the clinic ran beyond its planned finish.

Patient throughput

Measure: Number of patients seen in a given period.
Tool: **Number of patients seen per period**
Tool: number of patients seen in a given period (e.g. clinic / shift / period of time).

Financial impact & value for money

Hypothesis: The financial benefits gained by using AVT exceed its costs.

Return on Investment (ROI)

Measure: Financial return generated by AVT relative to its costs.
Tool: **Return on investment (ROI)** - consider costs (equipment, AVT licences, system integration, training, ongoing support) and benefits (time savings, increased patient throughput, cost savings). *Automated tools are not available for this.*

Payback period

Measure: Time required for AVT benefits to repay initial investment.
Tool: **Payback period** – calculated as the total initial investment (e.g. equipment, AVT licenses, training) divided by the annual net benefits. *Automated tools are not available.*

Measurement Tools

Tool	Used in AVT evaluation by
Word Error Rate (WER)	Imperial College London
Physician Documentation Quality Instrument (PDQI-9) tool	Nuffield Trust NIHR Rapid Service Evaluation Team (RSET) / Kernberg et al. (2024)
Sheffield Assessment Instrument for Letters (SAIL)	Great Ormond Street Hospital (GOSH) / Balloch et al. (2024)U.K. simulated pediatric (mock clinics)
The Consultation and Relational Empathy (CARE) Measure – a 10-item questionnaire to assess relational quality. The CARE Measure Website	CARE is well-validated in UK primary care but has not yet been adopted in AVT evaluation literature.
System Usability Scale Calculator	Shah et al. (2024). Stanford Health, U.S. ambulatory
Maslach Burnout Inventory Test	Central and North West London NHS Foundation Trust
Oldenburg Burnout Inventory	*Early Adopter*
Mini-Z Survey	Nguyen et al. (2023)Moffitt Cancer, U.S
NASA Task Load Index (NASA-TLX)	Great Ormond Street Hospital (GOSH)
The Net Promoter Score (NPS)	Great Ormond Street Hospital (GOSH)
TimeCaT	Great Ormond Street Hospital (GOSH)
Copenhagen Burnout Inventory	Central and North West London NHS Foundation Trust

Note: When selecting measurement tools, consideration should be given to licensing requirements, associated fees, attribution or citation requirements, permitted uses (e.g. research, clinical, or commercial), and any restrictions on modification or redistribution.

Evaluation guidance

- Determine which area(s) are significant for measurement (this Toolkit deliberately allows local discretion over *what* benefits are measured and *how* they are evidenced).

- Define which benefits the measures are intended to answer. Confirm what each metric indicates (output, outcome, or proxy for benefit).

- Collect baseline measurements prior to implementation and collect further measurements to be able to compare results to a baseline or appropriate comparator.

- Review change over time (short term, medium term and long term), not single data points.

- Combine metrics with qualitative insight (e.g. anecdotal data) to explain what is happening and why.

- Interpret findings to draw clear conclusions about service effectiveness, limitations, and learning. To do this, look at what was measured and move to what this tells us about the service.

References

- **“The Use of Ambient Voice Technology with Generative AI in Multiple Clinical Settings” (2025)** GOSH AVT Evaluation Report
- **National evaluation of Ambient Voice Technology in the NHS (Ongoing)** Nuffield Trust – Evaluation of AVT in the NHS
- **NHS England-backed AVT pilot evaluation (ED and outpatients)** Royal Devon leads NHSE-backed AVT pilot

- Balloch, J., Sridharan, S., Oldham, G., et al. (2024). Use of an ambient artificial intelligence tool to improve quality of clinical documentation. Future Healthcare Journal, 11, 100157. https://doi.org/10.1016/j.fhj.2024.100157
- Kernberg, A., Gold, J. A., & Mohan, V. (2024). Using ChatGPT-4 to create structured medical notes from audio recordings of physician–patient encounters: Comparative study. Journal of Medical Internet Research, 26, e54419. https://doi.org/10.2196/54419
- Nguyen, O. T., Turner, K., Charles, D., et al. (2023). Implementing digital scribes to reduce electronic health record documentation burden among cancer care clinicians: A mixed-methods pilot study. JCO Clinical Cancer Informatics, 7, e2200166. https://doi.org/10.1200/CCI.22.00166
- Shah, S. J., Devon-Sand, A., Ma, S. P., et al. (2025). Ambient artificial intelligence scribes: Physician burnout and perspectives on usability and documentation burden. Journal of the American Medical Informatics Association, 32, 375–380. https://doi.org/10.1093/jamia/ocae295



Health Innovation South West

NHS England South West

Thank You

@NHSSW

youtube.com/@nhssouthwest

england.nhs.uk/south