

=====

v1.0.0 24 August 2026 dataset synthetic-ris-v1

This census is generated from synthetic RIS assignment events. Figures are labelled illustrative. Protocol variance is the dispersion of protocol choice given the same indication, modality, and region - a finding about assignment, not a rating of appropriateness.

1. What protocol variance is

Protocol variance is the dispersion of protocol choice given the same indication x modality x region. Two protocolers can assign different protocols to the same ordered procedure; the cell records that fact as a count vector, not as a quality score.

The unit is the RIS assignment event. An appropriateness rating is not an input and is not an output.

2. Why it has never been measured

Claims files carry procedure codes, not protocol assignments. A CPT tells you what was billed, not which protocol was chosen for that indication.

Operations dashboards count exams completed. They do not partition those exams by protocol given the same indication cell.

A 24 August 2026 scan of public product pages found no named protocol-variance measurement feature from an operator, RIS vendor, or radiology-AI vendor. The gap is measurement, not another scanner.

3. The method

Knowledge object 7's first public half is the measurement, not a protocol selector. Four pieces are published here.

An indication cell is the triple indicationConcept x modality x region. Assignments are counted only inside one cell. Cross-cell pooling is refused.

The entropy statistic is normalised Shannon entropy $H / \ln(k)$ on the protocol-code count vector, bounded to $[0, 1]$. A single-protocol cell is exactly 0.

Cells with $n < 11$ return { suppressed: true, reason } and are never shown as a rate. Suppression is stated.

Grouping dimensions: protocoler (hashed), device (hashed), site, and shift. Groups are listed by hash, never ranked by entropy or share.

4. The illustrative result

The dataset is synthetic RIS assignment events (dataset synthetic-ris-v1). Each figure below carries the word illustrative on the same line as the number. It is not a customer figure.

[illustrative] Modal share, low-back-pain x mri x spine_lumbar: 56% n=32

[illustrative] Normalised entropy, low-back-pain x mri x spine_lumbar: 0.896 3 distinct protocols

[illustrative] Display string, low-back-pain x mri x spine_lumbar: modal protocol holds 56% of assignments across 3 distinct protocols in this indication cell

[illustrative] Modal share, headache x ct x head_brain: 67% n=24

[illustrative] Suppressed cell, abdominal-pain x ct x abdomen: suppressed (n=5; n < 11)

5. What a real extract would show

A live extract would fill the same cells with provenance measured rather than illustrative. Until that extract exists, these figures remain a worked example of the statistic, not a department's book.

The companion specification at `/api/publication/ris-extract` names the columns. It does not ask for pixels, reports, a patient name, a date of birth, an MRN, an address, or a clinician name.

6. What we do not claim

We do not claim that protocol variance is inappropriate care. We do not claim an improvement in appropriateness. We do not name or compare vendors. We do not treat this census as a product brochure.

This census does not claim that a protocoling product ships today. ARKA publishes the method and the extract specification so a department can run the same statistic on its own book.

7. How to run it on your own data

Send the RIS extract specified at `/api/publication/ris-extract`. Run `protocolVariance()` on the assignment events. Cells below n = 11 will suppress with a stated reason. Identical inputs plus the same engine pin produce a byte-identical result.

Extract specification: `/api/publication/ris-extract`

Census: `/api/publication/protocol-variance-census`