

Data Fitness for Use Checklist

Can this data set effectively answer your question(s)?

1. Defined Use Case

- ☐ What decision, task, or question will this data support?
- ☐ What level of precision is required (exploratory vs. clinical/regulatory)?
- ☐ What population, timeframe, or scope is needed?

2. Relevance

- ☐ Does the dataset contain the right variables for the task?
- ☐ Does it represent the correct population or context?
- ☐ Is the level of detail (aggregated vs. granular) appropriate?

3. Accuracy (Correctness)

- ☐ Do values correctly reflect real-world events or conditions?
- ☐ Are there known errors or inconsistencies?
- ☐ Has the data been validated against trusted sources?

4. Completeness

- ☐ Are all required fields populated?
- ☐ Are there significant missing values that could bias results?
- ☐ Does the dataset cover the full time period needed?

Data Fitness for Use Checklist

5. Consistency

- ☐ Are the data consistent over time and across systems and datasets?
- ☐ Are definitions, codes, and formats standardized?
- ☐ Are there conflicting values for the same entity?

6. Timeliness

- ☐ Is the data up to date for the intended use?
- ☐ Is it available when needed (real-time vs. delayed acceptable)?
- ☐ Does it reflect the correct time period?

7. Validity & Format

- ☐ Do the data conform to expected formats, rules, and standards?
- ☐ Are codes, units, and classifications valid?
- ☐ Are there out-of-range or illogical values?

8. Uniqueness & Integrity

- ☐ Are duplicate records present?
- ☐ If needed, can records be uniquely identified (e.g., patient ID)?
- ☐ Are relationships between data elements preserved?

9. Access & Provenance

- ☐ Are there barriers to use (permissions, delays, complexity)?
- ☐ Is the source trustworthy, well-governed, and fully documented?