Xerox – DESI V

Variability in TAR and Implications for Standards

Standards for Technology-Assisted Review

- **Standards for TAR are desirable and possibly essential for long-term viability.**

- **But which aspects of TAR should be standardized and how?**

- **Answering this question requires sensitivity to the many different sources and types of variability in TAR:**

  - Matter goals, budget, timeline, value, etc.
  - Intended application of TAR results (e.g. QC, prioritization, first-pass coding, etc.)
  - Observed TAR performance

TAR Performance Variability – Intrinsic Matter-Specific Factors

Checking intuitions:

- Richness
- Subject Matter
- Corpus Composition

TAR Performance Variability – Matter-Specific Execution Factors

Experimenting with supplemental inputs:

- Bigrams
- Metadata
- Pre-existing Models
- Multiple Supplementary Inputs

TAR Performance Variability – Matter-Specific Execution Factors

Facing reality:

- Counterevidence
- Time and Expense

Conclusion

There is no single formula for an optimal TAR project

- Matter-specific parameter tailoring is beneficial
- Ongoing algorithm innovation is healthy
- Imperfect TAR results may be perfectly fine
- Flexibility and adaptation are the keys to long-term TAR success

Standardize evaluation methodology, not process details

- Focus on best practices for model validation
- Provide guidelines for obtaining statistically sound performance metrics
- Let reasonableness, proportionality and the given use case dictate the details


xerox