DEM Research Summit List of Data Sources and Methods 2018-08-05 DRAFT

Potential Data Sources for Assessing Diagnostic Accuracy, Safety, or Quality

1. Voluntary reporting systems (morbidity & mortality conference logs, incident reports, patient complaints, non-mandatory patient registries)
2. General surveys of patients or providers (unknown denominator or low response rates)
3. Autopsies
4. Large ‘second review’ databases
5. Malpractice claims databases
6. Large administrative datasets (encounter/billing/utilization data, e.g., HCUP, Medicare LDS files, mandatory registries [e.g., state data on sexually transmitted diseases])
7. Targeted surveys of patients or providers (known denominator and high response rates)
8. Qualitative research findings from ethnography, focus groups, or related sources
9. Methodologically-rigorous, nationally-representative data samples (e.g., CDC’s NAMCS, NHAMCS, NHANES; AHRQ’s MEPS)
10. Systematic data collection in a prospective research study or structured QA/QI program
   a. direct observation during cognitive psychology (or related) experiments
   b. direct observation of clinical care (video or audio-taping, secret shoppers/standardized patients)
   c. random, periodic, or triggered clinical care audits (second reviews of chart, image/artifact, or patient)
   d. random, episodic, or triggered laboratory audits (calibration samples, rerun lab tests, repeat sample interpretation)
   e. systematic follow-up for test results or clinical outcomes (e.g., as part of a cohort study or diagnostic strategy RCT)

* Triggers are generally event-based (e.g., unplanned revisit within 14 days), but applied systematically
‘Ex vivo’ (Indirectly Clinical) Studies

1. Simulations & knowledge-based tests
2. Clinical practice surveys
3. Opinion surveys & focus groups
4. Consensus methods

‘In Vivo’ (Directly Clinical) Descriptive/Naturalistic Studies

5. Case report, case study, or case series
6. Systematic case audits
7. Case control (case-referent) study
8. Epidemiological/ecological descriptive or associative studies

Developmental (Early Stage) Diagnostic Accuracy & Error Studies

9. Methodological validation studies
10. Diagnostic technology development studies

Systematic, Prospective Non-Interventional Diagnostic Accuracy & Error Studies

11. Cross-sectional, paired (same patients/case material) diagnostic accuracy & error studies
12. Diagnostic cohort studies
13. Adjunctive qualitative studies married to a prospective, quantitative diagnostic study

Systematic, Prospective Interventional Studies

14. Quasi-experimental Intervention Studies
   a. Single-group ‘pre-post’ design
   b. Single-group time series design
   c. ‘Pre-post’ nonequivalent comparison group design
   d. Multiple-group time series design
   e. Stepped wedge
   f. Multi-group cluster randomized trial

15. Experimental Intervention Studies
   a. Control group RCT
   b. Solomon 4-group RCT
   c. Crossover RCT
   d. Crossover Solomon 4-group RCT

Synthetic Studies

16. Systematic review +/- meta-analysis
17. Diagnostic decision modeling studies
Potential Methods for Assessing Diagnostic Accuracy, Safety, or Quality

Diagnostic test studies are usually prospective, and retrospective diagnostic test studies should be viewed with some suspicion unless they are actually ambispective (i.e., nested within a prospective study of greater rigor). These studies may focus on diagnostic accuracy or impact (i.e., influence downstream diagnostic test use, alter clinical management, or affect patient outcomes).\(^1\) Accuracy studies are more common, but impact studies are more helpful for guiding clinical practice.\(^2\)

Diagnostic error studies are usually retrospective/descriptive, so suffer from the typical problems naturally associated with hindsight bias.\(^3\) Diagnostic errors can, however, be detected prospectively as part of a systematic study (e.g., physician diagnosis followed by immediate gold standard testing\(^4\)). Misdiagnosis-related harms are almost always determined retrospectively (due to ethical constraints), but can be studied prospectively in the context of a randomized trial of a diagnostic intervention.

‘Ex vivo’ (Indirectly Clinical) Studies

1. **Simulations & knowledge-based tests**
   a. Vignette studies (e.g., clinical reasoning,\(^5\) bias\(^6,7\)) & ‘secret shoppers’\(^8\)
   b. Knowledge-based tests (e.g., diagnostic misconceptions\(^9\))
   c. Biological decision-making metrics (e.g., eye-tracking,\(^10\) functional neuroimaging studies)

2. **Clinical practice surveys** (e.g., diagnostic approach\(^11\) or error\(^12\))

3. **Opinion surveys & focus groups** (e.g., brainstorming or defining problem space)

4. **Consensus methods** (for general or expert assessment of problem burden, causes, and solutions)
   a. Delphi method (for prediction) or nominal group technique (for prioritization)
   b. Crowdsourcing & social network analysis

‘In Vivo’ (Directly Clinical) Descriptive/Naturalistic Studies

5. **Case report, case study, or case series** (usually uncontrolled; sampling frame often unclear)
   a. Diagnostic error case feature descriptions\(^13-15\)*
   b. Process mapping,\(^16\) root cause analysis,\(^17\) or focused diagnostic error analysis\(^18\)*
   c. Other qualitative research study (ethnography, etc.)*

6. **Systematic case audits** (random, periodic, or targeted [autopsy,\(^19\) trigger-based\(^20\)] surveillance)

7. **Case control (case-referent\(^1\)) study** (non-nested, retrospective) (for rare disease or dangerous test)

8. **Epidemiological/ecological descriptive or associative studies** (these include large, population-based cross-sectional studies, time-trend analyses, and geographic variation studies with a well-described and systematically derived sampling frame; they often represent secondary analyses of ‘big data’)
   a. Disease prevalence studies (e.g., symptom prevalence, diseases ‘by symptom\(^21\))
   b. Diagnostic test resource use, diagnostic yield, and overdiagnosis studies (e.g., total utilization or costs or variation over time [time trend studies\(^22\)] or by region [small area variation studies\(^23\)], including inappropriate diagnostic test use\(^24\) and overdiagnosis\(^25\))
   c. Diagnostic error prevalence studies (general\(^26\) or symptom-specific,\(^27\) including diagnostic error proxy studies based on symptom-disease pair revisits\(^28,29\))

* Note that any of these data extraction methods may be used within the framework of a quantitative, prospective study, but they are more often seen applied in case reports or retrospective case series.
Developmental (Early Stage) Diagnostic Accuracy & Error Studies

9. Methodological validation studies (e.g., comparing methods of quality measurement\(^8,30\))
10. Diagnostic technology development studies (e.g., feasibility/usability, \(^31\) measurement validity\(^32\))

Systematic, Prospective Non-Interventional* Diagnostic Accuracy & Error Studies

Diagnostic studies include both screening (asymptomatic, demographic-based) and symptomatic testing protocols to detect diseases. New tests may be positioned as replacement tests, triage tests, or add-on tests (Figure below). Studies may be designed to focus on accuracy, precision, or reproducibility.\(^33\) The distinction between a cross-sectional design and a cohort design is not always straightforward for diagnostic studies. In general, however, in a cross-sectional design, the emphasis is on diagnostic accuracy at a point in time (even if the patients are followed over time for diagnostic confirmation based on future events), while in a cohort design, the emphasis is on prognosis and future event prediction.

11. Cross-sectional, paired (same patients/case material) diagnostic accuracy & error studies (single test, multiple test battery or diagnostic strategy, individual provider-based [provider ‘as test’])
   a. ‘Survey’\(^1\) approach (enrolls symptom or screens demographic population, tests all)
      ▪ Single test accuracy or precision relative to a reference standard test
      ▪ Comparative accuracy study (test A & B with reference to gold standard C\(^**\))
      ▪ Test battery (e.g., decision rule or algorithm development\(^34\) & validation\(^35\) studies)
   b. Concordance studies (inter-rater reliability, independent second reads)

12. Diagnostic cohort studies (defined by symptom, disease, or test result; patients followed over time)
   a. Prognostic indicator studies (e.g., diagnostic test assessing a risk factor\(^36\)) (Fig. 4.1, 4.5a)
   b. Repeated measures studies (test-retest reliability, trajectory studies)
   c. Nested diagnostic case-control study (ambispective) within a prospective cohort study
   d. Nested diagnostic test(s) cohort within a treatment RCT (test[s] done in all patients)
      ▪ Single test nested cohort (all patients randomized to treatment A vs. B) (Fig. 4.2)
      ▪ Multiple test nested cohort (Fig. 4.5)
         • All patients randomized to treatment A vs. B (Fig. 4.5b)
         • Discordant pair patients randomized to treatment A vs. B (Fig. 4.5c)

13. Adjunctive qualitative studies married to a prospective, quantitative diagnostic study
   a. Mixed-methods study (qualitative and quantitative planned together)
   b. Independent qualitative study

* Test results may be ‘concealed’ (non-interventional) or ‘revealed’ (quasi-non-interventional) as part of the study design; in the former case, incidental effects (e.g., Hawthorne-type) may influence clinical care; in the latter, clinical care may be affected, though it is not deliberately manipulated as part of the design.

** Note that in some instances, only a subset (e.g., discordant pairs only, when comparing tests A and B) may be subjected to the reference standard test to reduce expense or risk as part of the study design.
**Systematic, Prospective Interventional Studies** (for impact of new diagnostic test/strategy/pathway* or to reduce diagnostic errors/harm, including education, decision support, process-of-care interventions)

**14. Quasi-experimental Intervention Studies**
   a. Uncontrolled (single group) intervention studies
      - Single-group ‘pre-post’ design
      - Single-group time series design (multiple measures pre- and post-)
   b. Controlled (two-or-more-group), non-randomized intervention studies
      - ‘Pre-post’ nonequivalent comparison group design
      - Multiple-group time series design (multiple measures pre- and post-)
   c. Group** randomized studies (individuals within groups not randomly assigned)
      - Stepped wedge37 (staged rollout, order randomized, all receive intervention) – leverages staged rollout to control for epoch-specific confounders, increasing believability over a synchronous rollout; used when the group/center ‘n’ is too small for a properly-powered multi-group cluster RCT or when it is unethical for some groups to receive a placebo / non-intervention
      - Multi-group cluster randomized trial (half receive intervention)

**15. Experimental Intervention Studies** (individuals randomized)
   a. Controlled, non-equivalent arms (vs. placebo / standard care or vs. active control)
      - Control group RCT (‘post-only’ or ‘pre-post’-type measurement)
         - Test-results-based RCT (test done in all patients; only patients with a particular test result are randomized to treatment arms) (Figs. 4.3, 4.5c)
         - Randomize-to-reveal RCT (test done in all patients; patients randomized to results concealed or not, latter affecting clinical care) (Figs. 4.4, 4.6a)
         - Randomize-to-test RCT (diagnostic test, strategy, or decision support, done only in patients randomized to test protocol) (Fig. 4.4, 4.6b, 4.7)
      - Solomon 4-group RCT – 2x2 factorial with the intervention vs. control and, in each arm, half receiving the pre-test and post-test, the other half receiving only the post-test – used to segregate the effect of the intervention from the effect of repeated testing/outcome assessment, when there is concern about ‘test effects’ (i.e., measuring the ‘pre-’ may influence the ‘post-’ test) [typical for educational interventions or those with ‘paper-and-pencil’ outcome measures]
   b. Controlled, crossover (esp. useful if placebo /non-intervention would be unethical and the two possible interventions are not expected to interact substantially with one another; this is typical in the educational setting or with some safety interventions)
      - Crossover RCT (e.g., an educational or intervention where the two interventions have a slightly different focus and the interaction is anticipated to be minimal – one on diagnosing dizziness, the other on diagnosing headache; or one focused on metacognition and the other on communication skills)
      - Crossover Solomon 4-group RCT
* Interventions may be purely diagnostic, or be paired to explicit guidelines for further testing ('diagnostic strategies') or treatment (test-treatment combinations or 'management pathways')

** Groups (i.e., ‘clusters’) may be health systems, hospitals, departments/units/divisions, or any other meaningful unit (other than an individual) that can be randomized. In the case of randomizing individual clinicians, these should be analyzed as cluster-randomized designs when outcomes are related to groups of patients treated by that clinician (e.g., assessing the impact of a clinician training program on patient outcomes), and as an experimental (individual subject) design when each clinician is a research subject and unit of analysis (e.g., assessing the impact of clinician training on diagnostic knowledge or accuracy).

** Synthetic Studies (of diagnostic errors, tests, strategies, or interventions)

16. Systematic* review +/- meta-analysis
17. Diagnostic decision modeling studies
   a. Decision analysis
   b. Cost-effectiveness & cost-benefit analysis
   c. Estimated value of information analyses (‘meta-level’ effectiveness studies, answering questions such as “What is the anticipated return on investment for research on reducing diagnostic errors for this particular disease or disorder?”)

* Note that a narrative (i.e., unsystematic) review is not a research method.

---

Figure 1 (Bossuyt, BMJ 2006³⁸)
REFERENCES


