Today's Menu

Last Seminar: 2nd half planning

This Seminar: Systematic Literature Review (SLR)

Next Seminar: Discussion on SLR (Tan, Ken, Jon, & Sreekanth)

EBSE (Evidence-Based SE)

Key publications


Software engineers might make incorrect decisions about adopting new techniques if they don't consider scientific evidence about the techniques' efficacy. They should consider using procedures similar to ones developed for evidence-based medicine.

Evidence-Based Medicine

Dr: Problems (symptoms)?
Dr: Johnny has stomach flu.
Dad: How can you fix it?
Dr: He can take Drug A?
Dad: Wait a second. What's the evidence that A will work for him?

We all need information!

Hierarchy of Evidence in Medicine

Strongest

UK Centre for Reviews and Dissemination (CRD)
Guidance for Undertaking Reviews in Health Care

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Experimental studies, i.e., randomized controlled trial (RCT)</td>
</tr>
<tr>
<td>2</td>
<td>Quasi-experimental studies, i.e., studies without randomization</td>
</tr>
<tr>
<td>3a</td>
<td>Controlled observational studies</td>
</tr>
<tr>
<td>3b</td>
<td>Case control studies</td>
</tr>
<tr>
<td>4</td>
<td>Observational studies without control groups</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion based on theory, laboratory research or consensus</td>
</tr>
</tbody>
</table>

Weakest
Hierarchy of Evidence in Medicine (version 2)

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Evidence obtained from systematic reviews of all relevant randomized trials</td>
</tr>
<tr>
<td>II</td>
<td>Evidence obtained from at least one properly-designed randomized controlled trial</td>
</tr>
<tr>
<td>III-1</td>
<td>Evidence obtained from well-designed pseudo-randomized controlled trials (i.e., non-random allocation to treatment)</td>
</tr>
<tr>
<td>III-2</td>
<td>Evidence obtained from comparative studies with concurrent controls or interrupted time series with a control group</td>
</tr>
<tr>
<td>III-3</td>
<td>Evidence obtained from comparative studies with historical control or interrupted time series without a parallel control group</td>
</tr>
<tr>
<td>IV</td>
<td>Evidence obtained from case series, either post-test or pretest/post-test</td>
</tr>
</tbody>
</table>

Australian National Health and Medical Research Council (NHMRC) Guideline

5 Steps in Evidence-Based Medicine

<table>
<thead>
<tr>
<th>Step</th>
<th>Evidence-based Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Determining the need for information (about prevention, diagnosis, prognosis, therapy, causation, etc.) into an answerable question.</td>
</tr>
<tr>
<td>2</td>
<td>Tracking down the best evidence with which to answer that question.</td>
</tr>
<tr>
<td>3</td>
<td>Critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect), and applicability (usefulness in our clinical practice).</td>
</tr>
<tr>
<td>4</td>
<td>Integrating the critical appraisal with our clinical expertise and with our patient’s unique biology, values and circumstances.</td>
</tr>
<tr>
<td>5</td>
<td>Evaluating our effectiveness and efficiency in executing Steps 1-4 and seeking ways to improve them both for next time.</td>
</tr>
</tbody>
</table>

5 Steps in Evidence-Based SE

<table>
<thead>
<tr>
<th>Step</th>
<th>Evidence-based Software Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Determining the need for information (about development and maintenance methods, management procedures, etc.) into an answerable question.</td>
</tr>
<tr>
<td>2</td>
<td>Tracking down the best evidence with which to answer that question.</td>
</tr>
<tr>
<td>3</td>
<td>Critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect), and applicability (usefulness in software development practice).</td>
</tr>
<tr>
<td>4</td>
<td>Integrating the critical appraisal with our software engineering expertise and with our stakeholders’ values and circumstances.</td>
</tr>
<tr>
<td>5</td>
<td>Evaluating our effectiveness and efficiency in executing Steps 1-4 and seeking ways to improve them both for next time.</td>
</tr>
</tbody>
</table>

SLR (Systematic Literature Review)

- **Recommended EBSE method for aggregating evidence**
  - Identify the research questions
  - Define inclusion and exclusion criteria
  - Identify all relevant primary studies
  - Assess the quality of the primary studies

- **Quality in terms of what?**
  - Reporting versus Actual Study
  - Evidence: Is there any? Quantifiable?
  - Were the conclusions drawn in a convincing manner?

- **Practitioner focus**
  - Did the primary study answer practical questions?
  - Are mathematical estimating models more accurate than expert opinion based estimates?
  - What’s the level of overrun of SW projects and is it changing over time?
  - Is testing better than inspection?
Research Questions (to whom this SLR matters)

Tan
% How has the field of "Software Reuse Metrics and Models" been evolved since 1996?
  ➢ Is there new evidence? Is the new evidence supportive or contradictory?
  ➢ Are there any new metrics and models proposed?
  ➢ Are the six categories/buckets devised in [Frakes-CUSR96] still valid?
    Shall we change, merge, or remove some existing buckets?
  ➢ Who have kept refining/publishing their metrics and models?
  ➢ Who have been the leading researchers in this field since 1996?
  ...

Nash
% What’s out there for “semantic-enabled requirements traceability”?
  ???

Automatic and Manual Search

- ACM Digital Library
- Compendex
- IEEE Xplore
- ISI Web of Science
- Kluwer Online
- ScienceDirect – Elsevier
- SpringerLink
- Wiley Inter Science Journal Finder

[reuse OR reusability] AND [metric OR metrics OR measure]

From 1/1/1996 to 3/24/2010
Page #: 8

Deciding Primary Studies

Stage 1
Identify relevant studies – search databases and conference proceedings
n = 1,096

Stage 2
Exclude studies on the basis of titles
n = 621

Stage 3
Exclude studies on the basis of abstracts
n = 270

Stage 4
Obtain primary papers and critically appraise studies
n = 36
Data Extraction

Depending on your research questions

- How has the field on "Software Reuse Metrics and Models" been evolved since 1996?
  - Is there new evidence? Is the new evidence supportive or contradictory?
  - Are there any new metrics and models proposed?
  - Are the six categories/buckets devised in [Frakes-CUSR96] still valid?
  - Do we need new buckets? Shall we merge or remove some existing buckets?
  - Who have kept refining/publishing their metrics and models?
  - Who have been the leading researchers in this field since 1996?

Reporting

- Results (quantitative & qualitative)
  - E.g., in total, X new reuse metrics were proposed from 1996 to 2010; Y% of those were in Bucket #3 (amount of reuse); the progress of the field is slow (compared to 1980s); the quality of reported studies are higher (compared to 1980s).

- Trends & implications for research and practice
  - What's solved? (evidence)
  - What's open?
  - Available tools & success / failure stories?

Nonsystematic Literature Review

<table>
<thead>
<tr>
<th>Feature</th>
<th>Traditional reviews</th>
<th>Systematic reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Often broad in scope</td>
<td>Often a focused research question</td>
</tr>
<tr>
<td>Identification of research</td>
<td>Not usually specified, potentially biased</td>
<td>Comprehensive sources and explicit search strategy</td>
</tr>
<tr>
<td>Selection</td>
<td>Not usually specified, potentially biased</td>
<td>Criterion-based selection, uniformly applied</td>
</tr>
<tr>
<td>Appraisal</td>
<td>Variable</td>
<td>Rigorous critical appraisal</td>
</tr>
<tr>
<td>Synthesis</td>
<td>Often a qualitative summary</td>
<td>Qualitative and/or quantitative summary</td>
</tr>
<tr>
<td>Inferences</td>
<td>Sometimes evidence-based</td>
<td>Usually evidence-based</td>
</tr>
</tbody>
</table>
SLR & Reuse

They’re totally different!
SLR is about gathering evidence.
Reuse is about …
but there’s a fundamentally same aspect behind these two …

There must exist something (primary studies OR software artifacts) to be reviewed OR reused!
Can an SLR be a 1st level study?

Is everybody convinced?

We all need information!
No matter what level of study we do, people have stakes in it!