



KEELE
UNIVERSITY



Current State of Evidence-Based Software Engineering

Barbara Kitchenham

1

© Kitchenham 2007



KEELE
UNIVERSITY



Agenda

- Background
- Aims
- Method
- Results
- Conclusions

2

Background

- At ICSE04
 - Kitchenham, Dybå, and Jørgensen, proposed adopting Evidence-Based Software Engineering (EBSE)
 - Followed by papers at Metrics05 and in IEEE software
- As a result
 - Keele proposed a research project to investigate EBSE
 - Funded by EPSRC
 - For Keele & Durham
 - Now have a joint follow-on project (EPIC)

3

Evidence-based Practice

- Evidence-based Practice
 - Started in medicine
 - Expert opinion not as good as scientific evidence
 - Using best evidence saves lives
 - Being adopted/evaluated in many domains
 - Criminology
 - Social policy
 - Economics
 - Nursing
 - Management Science
 - Public health
 - Speech therapy

4

Goal of EBSE

- EBM: Integration of best research evidence with clinical expertise and patient values
- EBSE: Adapted from Evidence-Based Medicine
 - To provide the means by which current best evidence from research can be integrated with practical experience and human values in the decision making process regarding the development and maintenance of software
- Might provide
 - Common goals for research groups
 - Help for practitioners adopting new technologies
 - Means to improve dependability
 - Increase acceptability of software-intensive systems
 - Input to certification process

5

What is Evidence?

- Synthesis of best quality scientific studies on a specific topic
 - Main method
 - Systematic reviews
 - Methodologically rigorous synthesis of all available research relevant to a specific research question
 - Not ad hoc literature reviews
- Interpretation of research results to deliver guidelines for practitioners
- Consideration of research in specific contexts
 - Clients' Requirements
 - Current systems & expertise of staff

6

Practicing EBM & EBSE

- Sets requirements on practitioners and researchers
- Practitioners
 - Need to track down & use best evidence in context
- Researchers need to provide best evidence

7

EBSE Project

- Activities
 - Performing Systematic Literature reviews
 - Technology Acceptance Model
 - OO Design
 - Interviews with experts in other domains
 - Looking for experiences outside the medical domain to help revise guidelines
 - Compiling experiences of SLR process
 - Experiments with Structured Abstracts
 - Assessing status of EBSE

8

Aims and Method

- Aim
 - To present an overview of the current status of EBSE
- Method
 - A survey of papers addressing EBSE
 - Systematic Literature Reviews
 - Including Meta-analysis
 - Evidence-based guidelines for practitioners
 - Articles addressing EBSE
- Definitions
 - Primary studies are direct investigations of a topic or research question
 - Secondary studies (SLRs) synthesise primary studies
 - Tertiary studies synthesise secondary studies
- This is a tertiary study looking at research trends in SLRs
 - Following basic methodology of SLR

9

Research Question(s)

- How much EBSE activity has there been since 2004?
- What research topics are being addressed?
- Who is leading EBSE research?
- What are the limitations of current research?

10

Search Process

- Hand search of journals and conference papers since 2004
 - IST
 - JSS
 - IEEE TSE
 - IEEE Software
 - ISES05
 - ICSE04, 05 & 06
 - CACM
 - ACM Surveys
- Direct access to SIMULA & several researchers
- Still ongoing

11

Inclusion & Exclusion Criteria

- Include
 - Systematic Literature Reviews (SLRs)
 - Literature surveys with defined research questions, search process, data extraction and data presentation
 - Meta-analyses (MA)
 - Evidence-based practitioner guidelines (EBG)
- Exclude
 - Informal literature surveys (no defined research questions, no search process, no data extraction process)
 - Papers discussing process of EBSE

12

Quality Assessment

- DARE Criteria
 - Centre for Reviews and Dissemination (CDR)
Database of Abstracts of Reviews of Effects
- Questions
 - Are the review's inclusion and exclusion criteria described and appropriate ?
 - Is the literature search likely to have covered all relevant studies?
 - Did the reviewers assess the quality/validity of the included studies?
 - Were the basic data/studies adequately described?
- Answers: Yes (1), No (0), Partly (0.5)

13

Data Extraction

- Data required
 - Classification of paper
 - Type (SLR, MA, EBG)
 - Scope (Research trends or specific research question)
 - Main topic area
 - Research question/issue
 - Summary of papers
 - Quality evaluation
- Process
 - Extracted by one person
 - Checked by another person

14

Studies found

- 23 relevant studies
 - 1 meta analysis
 - 20 SLRS
 - 2 positioned as EBSE papers
 - 2 including evidence-based guidelines for practice
 - 2 EBG

15

Summary Results -1/3

- Scope
 - 9 of 20 SLR were research trends
- Topic
 - 9 papers on Cost estimation (including both EBGs)
 - 4 papers on Software Experiments
 - 3 papers on Testing
- Source
 - 17 papers had European authors
 - 4 had North America authors
 - 11 articles had authors from Simula Laboratory (Norway)

16

Summary Results – 2/3

- Sources
 - TSE: 4
 - IEEE SW: 4
 - IST: 3
 - JSS: 3
 - ICSE06: 1 (04 & 05 none)
 - ISESE05: 2
 - CACM: 1
 - ACM Surveys: 0

17

Summary Results – 3/3

- Quality of SLRs and MA
 - All papers scored 1 or more
 - One paper scored 4
 - Kitchenham, Mendes and Travassos
 - Systematic Review of Cross- vs. Within-Company Cost Estimation Studies, IEEE Trans on SE (short version published in EASE06).
 - Two papers scored 3.5
 - Magne Jørgensen
 - Estimation of Software Development Work Effort: Evidence on Expert Judgement and Formal Models, International Journal of Forecasting. (2007)
 - Zannier et al.
 - On the Success of Empirical Studies in the International Conference on Software Engineering. ICSE06
- Few papers performed a quality assessment
 - 3 fully & 4 partly

18

Specific Research Questions – 1/2

- Cost Estimation
 - Are mathematical estimating models more accurate than expert opinion based estimates?
 - No
 - What is the level of overrun of software projects and is it changing over time?
 - 30% and unchanging
 - Are regression-based estimation models more accurate than analogy-based models?
 - No
 - Should you use a benchmarking data base to construct an estimating model for a particular company if you have no data of your own?
 - Not if you work for a small company doing niche applications
 - Do researchers use cost estimation terms consistently and appropriately?
 - No they confuse prices, estimates, and budgets
 - When should you use expert opinion estimates?
 - Use expert opinion when you don't have a calibrated model or important contextual information is not available
- Cost estimation area also has Evidence-based Guidelines
 - No standards for constructing EPGs
 - No standard for evaluating their quality

19

Specific Research Questions – 2/2

- Testing
 - Is testing better than inspections.
 - Yes for design documents, No for code.
 - Which capture-recapture methods are used to predict the defects remaining after inspections?
 - Most studies recommend the Mh-JK model
 - Only one of 29 studies was an application study
 - What Empirical studies have addressed unit testing?
 - Empirical studies in unit testing are mapped to a framework and summarized.

20

Research Trends – 1/2

- Software Engineering experiments
 - How often do we do experiments in SE and what are their characteristics?
 - 103 out of 5453 articles searched
 - 33% on inspections
 - 66% tasks<2hours
 - 73% students
 - Do SE experiments consider theory and what sort?
 - 24 of 103 referred to theory
 - Is effect size reported in SE experiments and how large is it?
 - 29% of papers reported effect size.
 - Effect size was similar to psychology
 - What is the power of SE experiments?
 - Substantially below accepted norms (insufficient numbers of participants)

21

Research Trends – 2/2

- Others
 - What type of research is done in Computer science?
 - What type of research is done in Computer Science disciplines and how does it compare across disciplines (IS, SE, Computing)?
 - What type of evaluation studies are reported in ICSE?
 - What type of research is done in the area of Cost Estimation?
 - How rigorous is Web Science research?

22

Discussion – 1/5

- A relatively large proportion of SLRs relate to research trends
 - Disappointing since not of direct relevance to practitioner
 - SE experiment studies may have a long term effect
 - Improving empirical studies
 - Increasing reliability of basic evidence

23

Discussion – 2/5

- Simula Laboratory staff have made a significant contribution to EBSE
- Have adopted a useful strategy
 - Construct databases of primary studies related to research topics
 - Cost estimation
 - Software Experiments
 - Provide basic source material for many systematic literature reviews

24

Discussion – 3/5

- Quality is OK but could be improved
 - 16 of the 21 SLRs scored 2 or more
 - Few SLRs performed a quality assessment
 - Not important for papers covering research trends
 - Should be a critical part of a systematic literature review addressing specific research questions
 - Research trends papers don't need to report details of each paper
 - Score at best 0.5 on question 4
 - A simple way to improve scores against the DARE criteria is to report the search process
 - Papers that did not report their search process
 - Scored 0 for question 2 (effectiveness of search process)

25

Discussion – 4/5

- Cost estimation results demonstrate
 - EBSE can address practitioner related issues
 - Evidence can be used to develop practice-oriented guideline
 - However, no agreed method
 - For developing guidelines
 - Assessing their quality

26

Discussion – 5/5

- Testing results are a bit disappointing
 - Surprising that unit test search found only 24 primary studies
 - Compared with the study of capture-recapture model which found 29 experiments
 - A more extensive search process might deliver benefits
 - More studies
 - More specific research questions
 - Surprising that inspection results have not been subject to more formal evaluation
 - Narrative summaries have been published
 - No systematic literature review or meta-analysis
 - Feasibility study published but not followed up

27

References

- Barbara Kitchenham, Tore Dybå and Magne Jørgensen. (2004) Evidence-based Software Engineering. Proceedings of the 26th International Conference on Software Engineering, (ICSE '04), IEEE Computer Society, Washington DC, USA, pp 273 – 281 (ISBN 0-7695-2163-0)
- Tore Dybå, Barbara Kitchenham, and Magne Jørgensen. Evidence-based Software Engineering for Practitioners, IEEE Software, Volume 22 (1) January, 2005, pp58-65.
- Magne Jørgensen, Tore Dybå, and Barbara Kitchenham. Teaching Evidence-Based Software Engineering to University Students, 11th IEEE International Software Metrics Symposium (METRICS'05), 2005, p. 24.

28



Primary Studies

- Barcelos, R.F., and Travassos, G.H. (2006) Evaluation approaches for Software Architectural Documents: A systematic Review, Ibero-American Workshop on Requirements Engineering and Software Environments (IDEAS). La Plata, Argentina.
- Dyba, Tore; Kampenes, Vigdis By; Sjøberg, Dag I.K. (2006) A systematic review of statistical power in software engineering experiments, *Information and Software Technology*, 48(8), pp 745-755.
- Galin, D. and Avrahami, M. (2005) Do SQA programs work - CMM works. a meta analysis. *IEEE International Conference on Software - Science, Technology and Engineering*.
- Glass, Robert L., v. Ramesh and Iris Vessey. An Analysis of Research in Computing Disciplines CACM, 2004, 47(6), pp 89-94
- Grimstad, Stein, Jørgensen, Magne, and Molokken-Ostfold, Kjetil. (2006) Software effort estimation terminology: The tower of Babel, *Information and Software Technology*, 48 (4), pp 302-310.
- Hannay, Jo E., Dag I.K. Sjøberg, and Tore Dybå. A Systematic Review of Theory Use in Software Engineering Experiments. *IEEE Trans on SE*, 33 (2), 2007, pp 87-107.
- Jørgensen, M. (2004) A review of studies on expert estimation of software development effort, *Journal of Systems and Software*, 70 (1-2), pp37-60.
- Jørgensen, M. (2005a) Evidence-based Guidelines for Assessment of Software Development Cost Uncertainty, *IEEE Transactions on Software Engineering*, 31 (11) 942-954.
- Jørgensen, M. (2005b) Practical Guidelines for Expert-Judgment-Based Software effort estimation. *IEEE Software*, May/June, pp2-8.
- Jørgensen, M (2007) Estimation of Software Development Work Effort: Evidence on Expert Judgement and Formal Models, *International Journal of Forecasting*.
- Jørgensen, M., and Shepperd, M. (2007) A Systematic Review of Software Development Cost Estimation Studies, *IEEE Transactions on SE*, 33(1), pp33-53.



Primary Studies

- Juristo, N., A.M. Moreno, S. Vegas, M. Solari. (2006) In Search of What We Experimentally Know about Unit Testing, *IEEE Software*, 23 (6), pp72-80.
- Kampenes, Vigdis By, Tore Dybå; Jo E.Hannay; Dag I.K.Sjøberg. (2007) A systematic review of effect size in software engineering experiments. *Information and Software Technology*, In press.
- Kitchenham, B., Emilia Mendes, Guilherme H. Travassos (2007) A Systematic Review of Cross- vs. Within-Company Cost Estimation Studies, *IEEE Trans on SE* (short version published in EASE06).
- Mair, C. and Shepperd, M. (2005) The consistency of empirical comparisons of regression and analogy-based software project cost prediction, *International Symposium on Empirical Software Engineering. A systematic Review of Theory Use in Software Engineering Experiments*
- Mendes, E. (2005) A systematic review of Web engineering research. *International Symposium on Empirical Software Engineering*.
- Moløkken-Østfold, K.J.; M. Jørgensen; S.S. Taniikan.; H. Gallis.; A.C. Lien.; S.E. Hove. A Survey on Software Estimation in the Norwegian Industry, *Proceedings Software Metrics Symposium*, 2005.
- Petersson, H., Thelin, T, Runeson, P. and Wholin, C. Capture-recapture in software inspections after 10 years research – theory, evaluation and application, *JSS*, 72, 2004, pp 249-264
- Ramesh, V.; Glass, Robert L.; Vessey, Iris. (2004) Research in computer science: an empirical study, *Journal of Systems and Software*, 70(1-2), pp165-176.
- Runeson, P.; Andersson, C; Thelin, T; Andrews, A; Berling. What do we know about Defect Detection Methods? *IEEE Software*, 23(3) 2006, pp 82-86.
- Sjøberg, D.I.K.; Hannay, J.E.; Hansen, O.; Kampenes, V.B.; Karahasanovic, A.; Liborg, N.K.; Rekdal, A.C. A survey of controlled experiments in software engineering. *IEEE Transactions on SE*, 31 (9), 2005, pp733-753.
- Torchiano, M. Morisio, M. Overlooked Aspects of COTS-Based Development. *IEEE Software*, 2004.
- Zannier, Carmen, Grigori Melnick, and Frank Maurer. On the Success of Empirical Studies in the International Conference on Software Engineering, ICSE06, pp 341-350.