

SPI論文にみる再現性の問題例1

Takagi et al (2005) Emp Softw Eng 10, 495-515.

1. 終了した32PJを分類(12PJが失敗:Confusion)
2. PJM/PJLへのアンケート: PJ毎に開始時の準備不備状況22項目(要件定義、工数予測etc.)調査
3. 4項目を用いて失敗のリスク算出式を構成
4. 別の8PJ(2失敗、6成功)に適用し、7PJを的中

結果を見てからの記憶に頼るアンケートは偏る傾向
サンプルが小さく、統計学的に有意な関連ではない

統計学的にWell Writtenなので、Replication可能

SPI論文にみる再現性の問題例2

- 木田：SRAのプロセスデータを用いた
リスク解析（今日発表した内容）

全ての変数に値のあるPJのみ解析対象

どれかの変数に欠測値のあるPJは対象外

どれかの変数に欠測値のあるPJのリスク予測値
は過小評価の可能性が大きい

両論文とも統計的方法の記述は充分なので、
仮説として受け入れ、確認研究を実施すべき

“Evidence Based SPI” by Kitchenham 2003

SPI活動と品質との関連の確実な証拠を得る方法を、臨床医学に学ぶことを提唱

EBMによる、研究方法の信頼度ランク

1. **メタアナリシス**: 同一の仮説を検証した複数の研究報告の総合評価
2. **臨床試験**: Randomized Controlled Clinical Trial (無作為化調整介入実験)
3. **観察研究**: 自然経過を観察し、介入 (= 治療法) と背景因子と結果との関連を詳細に記述

無作為化臨床試験による薬効評価

無作為化

乱数で患者を2群に分割: 2群が対等を期待

対等な群

自然におけば評価変数の分布が等しい群
生存時間、検査値、**工数、不具合数** etc.

対等な群の一方のみに治療(介入)

有意な差 → 治療効果あり

偶然要因の影響は確率で評価: 検定

体質、環境、ケア、意志, etc.

臨床試験以前:チャールス2世の臨終

1685.2.2 am 8:00ベッドで突然けいれんし失神

14人の医師の治療記録

右腕血管から1Pint瀉血、肩を切り8オンス瀉血

催吐剤、下剤2回、浣腸2回

脳活性の為に頭を剃り頭皮に血豆、

野草の粉末、下剤、大麦湯、草・実の溶液を服用

舗装剤と鳩の糞を脚に塗る、瀉血

20以上の薬草、…でも悪化する。

けいれんを鎮めるのに人脳液40滴と

無数の動植物の抽出液を喉から注入、

牛王,果汁飲料,アンモニアを死体へ塗布

失敗の臨床試験：ジフテリア（カナダ）での実例

患者に無作為に新薬と旧薬を投与

追跡調査で死亡率比較

新薬群16% vs 旧薬群8%

新薬は旧薬より劣る。新薬は現在用いられている！

誤りの原因：医師達は

重症患者→新薬

軽症患者→旧薬

⇒無作為化を医師に任せ
ず乱数で実施すべき



SPI分野における論文の現状1

Sjøberg et al (2005)

IEEE Trans Software Eng 9

“A Survey of Controlled Experiments
in Software Engineering”

Software Engineeringにおける12の
一流 国際会議・雑誌に掲載された
5,453の論文を調査した結果

実験とみなされうるのは僅か103論文

論文の発表内容は不適切と結論

SPI論文の現状2

Dyba et al (2006) *Inf Softw Technol* 48

実験内容や統計解析法の理解が困難
Guidelineの必要性を強調

- Basili et al (2007)
Empirical Software Engineering 12,
“Protocols in the use of Empirical Software
Engineering Artifacts”

(1) *It is difficult to locate relevant information*

(2) *Important information is often missing.*

情報が容易に検索でき、方法を理解でき、結果を
正当に評価できるためのGuidelineを提案

- Jedlitschka and Pfahl (2005)
Int'l Symp. Empirical Software Engineering
“Reporting Guidelines for Controlled
Experiments in Software Engineering”
実験論文の書き方に関する提言と解説30頁

Jedlitschka and Ciolkowski (2005)
IESE-Report IASE-035.5/E

SPI実験結果の報告論文の内容と形式を提案

Altman et al (2001)による臨床試験の内容参照

Kitchenham et al (2008)
Empirical Software Engineering 13

提案Guideline の修正案44項目 数例を示すと

- Res-1 1 重要な環境要因と調整要因
- Res-4 7 検証する主仮説と副次的仮説
- Res-5 8 仮定している因果関係モデル
- Res-6 9 介入方法.
- Res-9 11 過去の研究との関連.
- Res-12 12 論文中で生データを公開しないなら、公開する条件
- Res-14 13 結果から結論を導いた根拠

SPIにおける介入実践案

仮説：要件定義の時間2倍==>仕様変更数は半分

無作為化実験

- **無作為化**：新規PJのうち10PJを無作為に選び介入
- **介入内容**：要件定義に予定の2倍時間をかける
- **統計解析**：2群間での仕様変更数の差のt-検定
- **判定**：差が有意のとき、仮説が正しい証拠

自然介入実験

- **自由意志**：介入を受け入れてくれる10PJに介入
- **背景要因**：全ての対象PJの属性データの収集
- **統計解析**：2群間での背景要因の違いを調整した上での仕様変更数の差の検定のための多変量解析
- **判定**：差が有意のとき、仮説を仮承認し調査継続

提案

1. Center for Defect Control 設立
JASPIC または 各社毎
Biostatisticianを調査委員に加える

2. SPI活動の成果報告は再現性を
確認できる内容にする

3. Journal of SPI
in Japan:
再現性確保のため
Biostatistical
Reviewer

Count
on
Biosta
tistics

