Di 9

Software Cost Overruns – How Large Are They and How Should They be Measured? A critique of the Standish Group Chaos report.

Magne Jørgensen

Organized by:

SIGS DATACOM

Lindlaustr. 2c, 53842 Troisdorf,
Tel.: +49 (0)2241 2341-100, Fax.: +49 (0)2241 2341-199
www.oopconference.com

[simula . research laboratory]

Software Cost Overruns - How Large Are They and How Should They be Measured?

Magne Jørgensen
Simula Research Laboratory

"The software industry is our laboratory"

Who am I?

- 10 years industry experience from Telenor (telecom) and Storebrand (insurance).
- Consultancy work for several Norwegian Companies.
- Position as professor at Simula Research Laboratory and University of Oslo.
- More than 20 journal papers on software cost estimation.

Overoptimistic predictions of own ability is human …

… the same goes for underestimation of risk

The Standish Group has tried to measure the level of over-optimism …

"The Standish Group research shows a staggering 31.1% of projects will be canceled before they ever get completed. Further results indicate 52.7% of projects will cost 189% of their original estimates." (1994 CHAOS Report)

*Can we trust these, frequently quoted, results? Are IT-people **extremely** over-optimistic?*

The Standish Group's 1994, 1996, 1998, 2000 and 2002 CHAOS Reports

- 1994: 189% average cost overrun (of challenged projects)
- 1996: 142% average cost overrun (of challenged projects)
- 1998: 69% average cost overrun (of challenged projects)
- 2000: 45% average cost overrun (of challenged projects)
- 2002: 43% average cost overrun (of challenged projects)

Is it possible that the software industry has improved so much in less than 10 years?

Or, could it be that the study of the Standish Group is seriously flawed?

PROBLEM 1: Lack of Precision

- What does 189% average cost overrun mean?
- Formulations used by the Standish Group in various reports and press releases:
  - "… *52.7% of projects* (the so-called challenged projects) *will overrun their initial cost estimates by 189%*"
  - "*The average cost overruns for combined challenged and cancelled projects is 189%.*"
  - "… *52.7% of projects will cost 189% of their original estimates*"

➔ 89% or 189% overrun?

➔ challenged, or challenged + cancelled projects?

➔ Estimate = most likely effort, planned effort, budget, price, ...?

Were we the only one to be confused?

NO! A web-search on the use of the CHAOS numbers gave the following results:

189% or 89% overrun?

- 50% of the documents described the result as "*189% cost overrun*",
- 40% as "*189% of original estimate*", and
- 10% as "*89% cost overrun*".

All projects, challenged projects, or challenged + cancelled projects?

- 70% of the documents related the result to "*53% of the projects*" (without explicitly pointing out that this 53% referred to challenged projects only),
- 16% to "*all projects*",
- 8% to "*challenged and cancelled projects*",
- only 6% explicitly pointed out that the average cost overrun is based on "*challenged projects*" only.

PROBLEM 2: Lack of correspondence with other surveys

Study	Jenkins [9]	Phan [10]	Bergeron [11]
Year	1984	1988	1992
Respondents	23 software organizations	191 software projects	89 software projects
Country of Respondents	USA	USA	Canada
Average Cost Overrun	34%	33%	33%

PROBLEM 3: Lack of description about research method

- No description of **accuracy measures (!!!)**, selection process, or analysis method.
- We have sent several mails to them and received responses like:
  - *"Why would you be surprised that we would not want to give out detailed information about how we conduct our studies? This is how we make a living. There is absolutely no incentive for us to want anyone to replicate our study - that's like giving away our business for free."*
- After three mails they stopped responding.
  - They were not even willing to respond on whether the 1994-study results should be interpreted as "89%" or "189%" cost overrun.
- Then, we tried to buy the (rather expensive) CHAOS report, but they didn't respond on that either, ...
  - We had to order the report through another company.

Possible reasons for the high cost accuracy numbers in the 1994 report

- Non-random sampling?
  - (From research report) *"We then called and mailed a number of confidential surveys to a random sample of top IT executives, asking them to share failure stories. During September and October of that year, we collected the majority of the 365 surveys we needed to publish the CHAOS research."*
- Confused by own imprecise definitions?
  - Seems as if the initial results were interpreted as 189% of initial estimate, i.e., 89% cost overrun. Later, however, this is by the Standish Group described as 189% cost overrun.

What can we learn from this?

- The Standish Group's results are probably flawed.
- Measurement of cost overruns is difficult.
- Research not subject to peer-reviews may be difficult to evaluate
  - No independent review of research quality
  - Poor research method hidden as "business secret"
- The number of participating companies is not a good indicator of research quality if the sample is biased, i.e., non-random.
- If something seems to be very wrong (e.g., an improvement from 189% to 45% cost overrun in 6 years) it probably is.
- **Control question:**
  - **Why did we believe in the 189% cost overrun number (and the corresponding 'software crisis') in the first place?**

[simula . research laboratory]

How should we measure cost overrun?

Suggested steps

1. Define the factor you want to evaluate through cost estimation accuracy measurement
2. Ensure a sufficiently precise definition of estimation accuracy (including a precise definition of 'estimate')
3. Decide how to isolate the factor you want to evaluate
4. Collect the information necessary for isolation
5. Analyze the factor to be studied

1) What do you want to measure?

- Do you want to evaluate:
  - Estimation ability?
  - Project cost control ability?
  - Project complexity?
- All these factors (and many more) impact the estimation accuracy.
- If you don't have a method for isolation, you will not know what you measure.

2) Definitions

- What is an effort (or cost) estimate?
  - Most likely/median/mean/most optimistic effort?
  - Planned effort?
    - Risk-averse planned effort?
    - Optimistically planned effort?
  - Budget? (e.g., planned effort + a contingency buffer)
  - Price?
- Surveys of estimation accuracy and most software organizations typically mix these interpretations.
  - The basis of cost estimation accuracy measurement is consequently very poor.

3) Isolation strategy

- Decide on strategy for isolation of the factor (e.g., the estimation ability factor) you want to measure.
- Possible isolation strategies include:
  - **Adjust** for the contribution of the non studied factors, e.g, increase in functionality from estimated to actual product.
  - **Divide** the projects into subsets with characteristics similar regarding the non studied factors.
  - **Exclude** observations from the analysis when non studied factors have a large but unknown impact on estimation error or the observation is not relevant for the goal of the estimation error measurement and analysis.
  - **Randomize** the treatment.

4) Collect information (Example)

To evaluate effect of estimation method on estimation accuracy an organization may have to collect information about:

- Difference between required functionality and quality at time of estimation and of the proportion actually delivered.
- Estimation complexity. (Especially, systematic difference projects where estimation methods are applied.)
- Flexibility of requirement specification, as assessed by the project manager. (Is estimation accuracy caused by flexibility?)
- Description of reasons for low and high estimation error (to exclude unusual projects).

5) Analyze (Example)

A (near real-life) example of what may happen if the data collection had not enabled a division of projects into projects of different estimation complexity.

- A function points-based estimation method was mainly applied on projects with low estimation complexity.
- The original finding was that the function point-based method was better.
- When separating into project of similar estimation complexity, however, the difference was gone!

Conclusion

Measurement without knowing

- What we measure
- Why we measure
- How to isolate the factors we want to measure

is waste of time.

This is, however, what we have been doing, and still do, in software cost accuracy measurement.

Documentation

- **How Large Are Software Cost Overruns? Critical Comments on the Standish Group's CHAOS Reports**
  - **To be published in Information and Software Technology (a preliminary copy can be downloaded from www.simula.no)**
- **A Framework for the Analysis of Software Cost**
  - **Submitted to Information and Software Technology, 2005 (a preliminary copy can be downloaded from www.simula.no)**