The rise and fall of the Chaos report figures

J.L. Eveleens and C. Verhoef

VU University Amsterdam, Department of Computer Science,
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

`laurenz@few.vu.nl, x@cs.vu.nl`

December 17, 2008

Abstract

In this paper we discussed the project success figures reported on by Standish Group. In 1994 Standish published the Chaos report that showed a shocking 16% project success. This and renewed figures by Standish are often used to indicate information technology is in a troublesome state. However, some authors have raised questions about the validity of these figures. In this paper we showed that Standish's successful and challenged project results are indeed meaningless for benchmarking. Our research on 12187 forecasts of 1741 real-world projects of in total 1059 million euro showed that IT forecasts have political biases. Yet, Standish's numbers, which depend on IT forecasts, do not account for these biases. Therefore, these figures have very limited use for benchmarking as we demonstrated using our real-world data.

Keywords and Phrases: CHAOS report, Standish group, forecasting, project success

1 Introduction

For many years researchers and practitioners are analyzing how to successfully manage IT projects. Among them is the Standish Group, which regularly publishes its findings in their so-called Chaos reports. In 1994, Standish reported a shocking 16% project success rate, another 53% of the projects had overruns of costs or time or less functionality and 31% of the projects failed outright. In subsequent reports Standish updated their findings, yet the figures remained troublesome. These reports, derived from their longitudinal data, suggest that many efforts and best practices to improve project management hardly help to increase project success.

Over the years their figures have had tremendous attention. Many authors have quoted the Standish figures to illustrate information technology is in a troublesome state. However, some authors, e.g. Glass and Jørgensen et. al. [8, 9, 17], have raised questions about the validity of the Standish figures, but have been unable to unquestionable refute their credibility. In this paper, we will show their reservations are justified. Based on our own research on 12187 forecasts of 1741 real-world projects of in total 1059 million euro, we found that the political bias of forecasts, e.g. forecasts of costs are made positively to make the project look good, is highly influential for their successful and challenged project figures. This, among others, causes these figures to be very limited in use for benchmarking.

2 Chaos reports

In 1994 Standish Group published the first in a series of reports, called the Chaos report [12]. This report is a summary of research findings of Standish Group, which aimed to investigate causes of software project failure and to find the key ingredients that are able to reduce such failures. Moreover, the research tried to identify the scope of software project failures. To achieve this latter goal, Standish defined three project categories, namely:

- A project is successful if it is completed on time, within budget, and with all features and functions as initially specified.
- A project is challenged in case it is completed and operational but over budget, over the initial time forecast, and if it offers fewer features and functions than originally specified.
- A project is impaired or failed if it is canceled at some point during the development cycle.

To find answers to their research questions Standish sent out questionnaires and received 365 respondents representing 8380 applications. Based on the responses Standish published overall percentages for each of the above categories in the Chaos report. In their first report they found a 16% project success rate, 53% challenged projects and 31% failed projects. Standish updated their figures in subsequent years as shown in Table 1. These figures were published in various white papers [10, 12, 13, 14].

Table 1: Standish project benchmarks over the years.

year	success	challenged	failed
1994	16%	53%	31%
1996	27%	33%	40%
1998	26%	46%	28%
2000	28%	49%	23%
2004	29%	53%	18%

The figures indicate large problems with software engineering projects, and as such have had an enormous impact on information technology. They suggest that the many efforts and best-practices to improve the way we develop software seem hardly successful. The numbers have been widely cited in scientific articles, for instance [24, 27, 28, 26, 1, 2, 4, 20, 11, 25], and the media, for instance [21, 3, 6, 22, 23]. Many authors use the figures to point out that software development is in a crisis. The numbers even found their way to the PITAC (President's Information Technology Advisory Committee) report [18], to substantiate the claim that both software products and processes in the United States are inadequate.

The impact of the figures and their widespread use, indicate that they are perceived to be impeccable and unquestionable. However, we will show that the successful and challenged project figures are unreliable for benchmarking purposes and as such have often been misused. At first sight, the only way to assess the credibility of the Standish results is to use their data and reiterate their analyzes. Other authors, e.g. Glass [8, 9] and Jørgensen et al. [16] also indicated that. But there is another method. This method is to obtain your own data and carry out comparable research in order to see whether their results are reproducible. By obtaining large amounts of data of four organizations, we were able to pinpoint an important aspect overlooked by Standish. This aspect, the bias of forecasts, is of such an influence that without taking it into account any aggregation of successful and challenged project figures of different organizations becomes meaningless. To properly explain this, we must first discuss how to analyze forecasting quality.

3 Forecasting quality

The Chaos report definitions are all about forecasts. Standish defines a project to be successful when the initial forecasts of costs and duration are higher than the actual costs and durations, and initial forecasts of functionality are lower than the actual functionality. But are these initial forecasts really accurate?

In [7] we performed an extensive study on 1741 real-world projects of in total 1059 million euro to, among others, answer this question. In 1981 Barry Boehm [5] showed that the quality of forecasts improves as a project progresses. Boehm showed this by drawing his now famous *cone of uncertainty*, a plot which depicts forecast to actual (f/a) ratios against project progression. We found that plotting these f/a ratios is very insightful to assess the forecasting quality. This plot gives information on how the forecasts are made, what uncertainty they contain and whether they are politically biased. In [7] we applied the method to data we received from four organizations, of which we will use three in this paper to prove our point. Below we will give a short summary of the findings for three of the four case studies we published. More details are discussed in [7].

Landmark Graphics Corporation Landmark Graphics is a vendor of commercial software that is used for oil and gas exploration and the production market. The data that we obtained from Todd Little, which he reported on in IEEE Software [19], consists of 121 software development projects carried out in the period of 1999 to 2002. In total Little provided us with 923 distinct forecasts that predict the duration of these 121 projects.

We analyzed the data by dividing the forecasted durations with the actual project durations and plotted the ratios as shown in Figure 1. The horizontal axis of the figure represents project progression. The start of a project is depicted at zero and project completion is represented by one. The vertical axis shows the value of the f/a ratio. For instance, the data point at project completion 0.2 and f/a ratio 0.5 indicates a forecast was made when the project was completed for one fifth. This forecast was half the actual, meaning the project turned out twice as expensive as anticipated.

The figure shows that most forecasts made in this organization are lower than the actual. This means that projects appear to take longer than initially anticipated. We found that the goal of this organization is to forecast the minimum value instead of the actual value. This caused most forecasts to be lower than the actual values.

Organization X We performed the same analysis on data from a large multinational organization. The data consists of 867 IT-intensive projects, which were all started and completed in 2005 or in 2006. In total 3767 forecasts of the costs of these projects were made. Figure 2 depicts the f/a ratios of this data.

This figure shows that the forecasts made in this organization are in general higher than the actual. Also, the data does not show a conical shape as we would expect

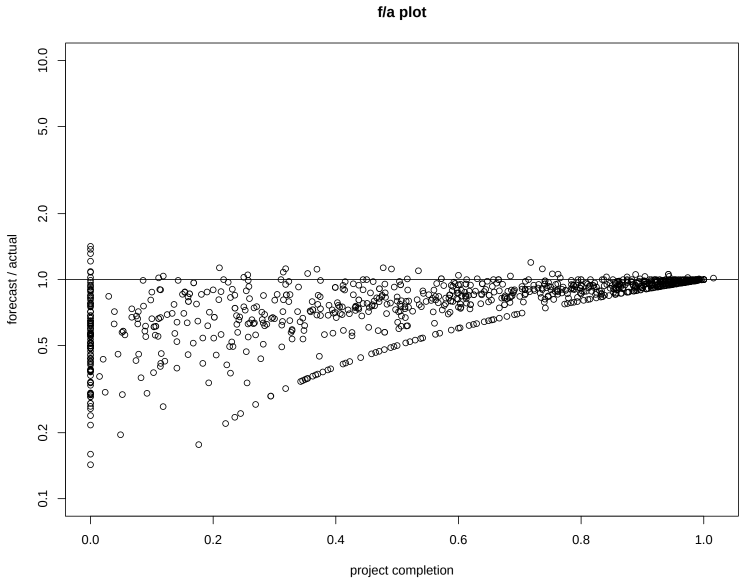

Figure 1: $923 f/a$ ratios for 121 project durations of Landmark Graphics.

from Boehm's *cone of uncertainty*. Projects usually had surplus budget after they were completed. After discussion with the organization, we found the forecasts were aimed at predicting the maximum value needed to complete the project or even more. Usually less funds were granted to a project than was requested. Also, obtaining additional funds later on in the project was difficult. Therefore, to account for this cultural aspect, requests for budgets were made considerably higher than the estimators thought the actual costs would be. This created enough buffer for projects to be carried out without running out of funds.

Furthermore, this organization steered on Standish project success indicators. They adopted the Standish definition to define when projects were successful. This also caused project managers to overstate budget requests to increase their safety margin of success.

Organization Y The third case study discusses a large multinational financial service provider. From this organization, we obtained data on 140 software development projects conducted in the period of 2004 to 2006. In total 667 forecasts were made for the total costs of these projects.

The f/a ratios in Figure 3 resemble Boehm's conical shape with the forecasts centered around the actual value. It turned out that in this organization the forecasts were assessed by an independent metrics group. This group made its own cost calculations next to that of project managers. If large discrepancies arose these needed to be resolved before any budget was approved. This caused forecasts to aim at predicting the actual value.

In two of the above case studies the forecasts are highly influenced by politics. In one organization most forecasts are lower than the actual values since they predict the

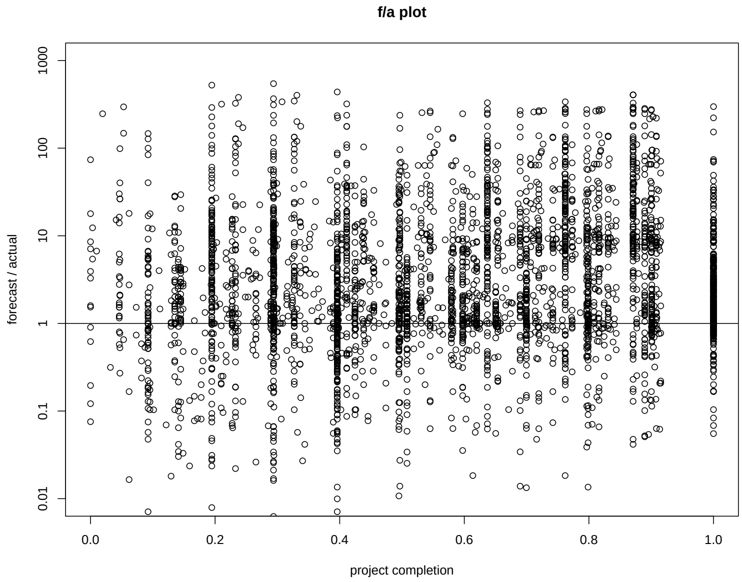

Figure 2: $3767 f/a$ ratios for 867 project costs of organization X.

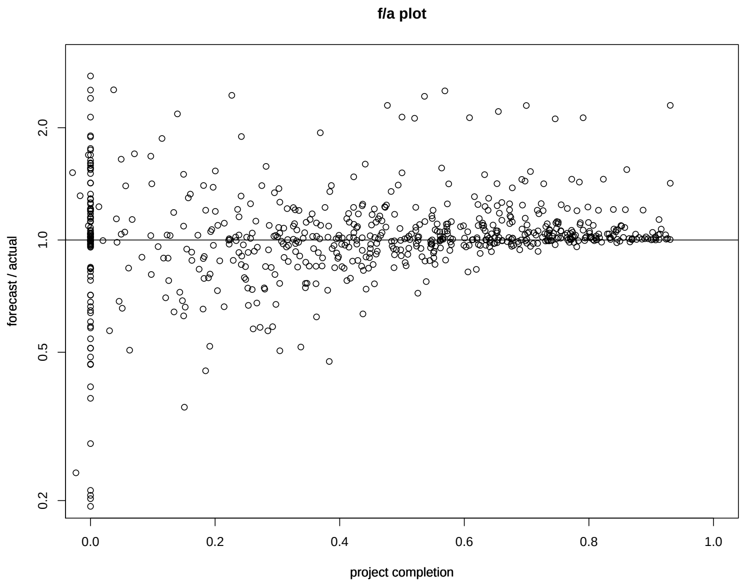

Figure 3: $667 f/a$ ratios for 140 project costs of organization Y.

minimal time required to finish the project. In another organization the forecasts are much higher than the actual values as large safety margins are taken into account. In the third organization we found no bias towards either under- or overestimation. But how do such politics affect the Chaos report figures? In the next section we will show in detail that biases of forecasts significantly influence these figures.

4 Applying Standish's definitions

To illustrate how the forecasting bias affects the Chaos report figures, we applied Standish's definitions to our case studies. Since these definitions are about the initial forecasts, we applied them only to the first forecast of each project. Note that this is a subset of all data points shown in the figures above. For each organization, we counted the number of projects that had a forecast larger than the actual, which indicates a successful project. We divided this by the total number of projects.

Note that the resulting figures are an upper bound (lower bound) for the successful (challenged) project Chaos figures. First, in each of the case studies we obtained data for only costs or time. If all three dimensions (cost, time and functionality) of these organizations are taken into account, the numbers will always be equal or lower than the successful percentages calculated for only one dimension. Therefore, they are an upper bound for these figures of each organization.

Second, the figures do not incorporate failed projects. If failed projects would be taken into account, again the numbers will always be equal or lower than the current successful percentages. However, since we focus on the successful and challenged project figures, the failed projects are not relevant for our purpose. Therefore, the figures are sufficient to prove that the successful and challenged figures are highly influenced by the biases of forecasts.

Table 2: Successful and challenged projects in our case studies.

source	success	challenged
Landmark Graphics	5.8%	94.2%
Organization Y	59%	41%
Organization X	67%	33%
1/Landmark Graphics	94.2%	5.8%

Table 2 shows the Standish percentages for our three case studies. The table provides an interesting insight in the Standish figures. It shows that organization X is very successful compared to the other case studies. Nearly 70% of the projects is successful according to the definitions of Standish. On the other end Landmark Graphics has only a 6% success rate. Yet, the case studies show that this is far from reality. The Landmark Graphics initial forecasts deviate far less from their actuals than in the case of organization X. For, that organization overestimates tenfold times up to hundredfold times as shown in Figure 2. Obviously, Landmark Graphics is more accurate in predicting the actual value than organization X. Organization X merely appears successful as their bias is to overestimate, which is an advantage in Standish's definition of success.

The case studies show that the Standish figures for individual organizations are highly influenced by forecasting biases. One may wonder whether the biases cancel each other out when averaging these numbers. However, it is bad practice to process erroneous numbers that do not represent reality, in order to reach a proper result. When the underlying data has an unknown bias, any aggregation of that data is unreliable and useless.

To show that it is easy to influence the Standish figures we added 1/Landmark Graphics. This fictional organization resembles the exact opposite of Landmark Graphics. That is, the deviations to the actuals remain the same, but an overrun becomes an underrun and vice versa. The resulting percentage becomes 94% making 1/Landmark Graphics highly successful. Thus, by overstating instead of understating their forecasts with similar accuracy, Landmark Graphics would suddenly become very successful according to Standish.

The influence of overstating forecasts on the Standish figures is not only evident from our figures. Standish, by means of their chairman Jim Johnson, clearly indicates it is easy to manipulate their figures by saying in [15]:

> In 1998, they [the estimators] had changed their [estimating] process so that they were then taking their best estimate, and then doubling it and adding half again.

Johnson made this statement with respect to the drop in their reported average cost overruns between 1996 (142%) and 1998 (69%). In the paper Johnson continues that he does not believe this change of process is the cause of the drop. However, our case studies show that forecasting biases have a large influence on such figures. Therefore, we believe the drop in the reported cost overruns is most likely caused by this change in the estimating process.

In [7], we developed methods that account for forecasting biases and prevent such influences. In that paper we used bandwidths surrounding the actual value to determine when underruns or overruns are successful and when not. With these bandwidths, projects with small under- or overruns, e.g. one dollar, are considered successful, whereas projects with large under- or overruns, e.g. a million dollar, become challenged projects. Using our methods we were able to derive figures that were exactly in line with the reality of our case studies. For more information we refer to [7].

Our case studies clearly show that by ignoring the potential bias and quality of the forecast, the figures of Standish Group do not give an adequate indication what, according to their definitions, constitutes a successful or challenged project. Since some organizations have the tendency to overestimate and others to underestimate, the successful and challenged project figures are in fact useless. Without knowing and accounting for these biases, the numbers are meaningless for benchmarking.

5 Discussion

This paper is not the first to challenge the credibility of the Chaos report figures. A number of authors also "questioned the unquestionable" [8] and raised concerns about the validity of the figures [9, 17]. For instance, Glass [8, 9] felt the figures do not represent reality. Without plentiful successful projects, he asserts, it is not possible to live in what some described as the current computer era. Moreover, he states that other objective research findings, in general, do not support the findings of Standish.

Also, Jørgensen et al. express their doubt about the numbers in their paper [17]. The paper unveils a number of issues with the definitions and argues the resulting figures are therefore unusable. One of the issues is that the definition of project success and project challenged focus on overruns and discard underruns. Although not stated by Jorgensen et al., we showed that neglecting underruns causes forecasting biases to highly influence the figures. Therefore, the one-sided approach of the definitions indeed has large consequences on the validity of Standish's figures.

Despite the valid questions raised by these authors, to date no-one was able to definitely refute the credibility of the Standish figures. By showing that the forecasting biases highly influence the successful and challenged project figures, we clearly illustrate that these figures are indeed unreliable for benchmarking.

Note that the we only addressed the success and challenged categories. Since forecasting bias is not directly connected to failing projects, we have not assessed those figures. One could use the Standish figures as failed projects and non-failed projects. In that case, the non-failed project category is the sum of successful and challenged projects.

6 Conclusion and remarks

In this paper we discussed figures published by Standish Group, which show the percentage of projects that succeed, are challenged or fail. These figures have risen to be some of the most famous and widely cited industry benchmarks in information technology. They are used up to the highest governmental levels to support the view that software development is in a crisis.

A number of authors questioned the usability of Standish's figures for benchmarking, but this did not lead to abandoning their results. To that end, we showed based on real-world data that Standish's definitions do not incorporate the bias of forecasts, which significantly influences the percentages of successful and challenged projects. Since their definitions heavily rely on the quality of forecasts, the definitions should account for the potential biases of predictions. We showed that in one organization forecasts were used to predict the minimum value, which resulted in forecasts that were mostly lower than the actual. Another organization generally overstated its cost estimations to increase its margin of error. Although successful in Standish's view, this caused forecasts to bear hardly any relation to the actual. A third organization had no bias towards under- or overestimating at all.

Without taking such biases into account the successful and challenged project figures are easily influenced without actually changing or improving anything. Therefore, these figures are meaningless for benchmarking and of insufficient quality to support any conclusion about the state of software development.

We communicated our findings with Standish Group, who responded by means of their chairman Jim Johnson. He wrote: "All data and information in the CHAOS reports and all Standish reports should be considered Standish opinion and the reader bears all risk in the use of this opinion." We fully support this disclaimer, which to our knowledge was never stated in the Chaos reports. We advice against using these Standish figures.

In our research we derived alternative definitions and methods that do incorporate the bias of forecasts. We will not address this here as it requires an elaborate and detailed discussion. For more information about our proposed definitions and methods we refer to our paper [7]. Finally, we hope that Standish will adopt our definitions and methods for the rise and resurrection of their reports.

References

[1] Ken Abernethy, George Piegari, and Han Reichgelt. Teaching project management: an experiential approach. *Journal of Computing Sciences in Colleges*, 22(3):198–205, 2007.

[2] Frank T. Anbari. Innovation, project management, and six sigma method. In M. Afzalur Rahim and Robert T. Golembiewski, editors, *Current topics in management*, pages 106–107. Transaction Publishers, 2005.

[3] Andrew Birmingham. Project perils. The Australian, May 4 2004.

[4] Kurt Bittner and Ian Spence. *Managing iterative software development projects*. Addison-Wesley, 2006.

[5] Barry Boehm. *Software engineering economics*. Prentice hall PTR, 1981.

[6] Michael Cross. Sauce for the goose. The Guardian, July 24 2003.

[7] J.L. Eveleens and C. Verhoef. Quantifying IT forecast quality. Under review. Available via www.cs.vu.nl/~x/cone/cone.pdf.

[8] Robert Glass. IT failure rates - 70% or 10-15%. *IEEE Software*, May:110–112, 2005.

[9] Robert Glass. The Standish report: does it really describe a software crisis. *Communications of the ACM*, 49(8), 2006.

[10] Deborah Hartmann. Interview: Jim Johnson of the Standish Group, 2006. www.infoq.com/articles/Interview-Johnson-Standish-CHAOS.

[11] Kathleen B. Hass. *Professionalizing business analysis*. Management concepts, 2007.

[12] The Standish Group International Inc. Chaos. Technical report, The Standish Group International Inc., 1994.

[13] The Standish Group International Inc. Chaos: A recipe for success. Technical report, The Standish Group International Inc., 1999.

[14] The Standish Group International Inc. Extreme chaos. Technical report, The Standish Group International Inc., 2001.

[15] Jim Johnson. Standish: Why were project failures up and cost overruns down in 1998, 2006. www.infoq.com/articles/chaos-1998-failure-stats.

[16] Magne Jørgensen. Experience with the accuracy of software maintenance task effort prediction models. *IEEE Transactions on software engineering*, 21:674–681, 1995.

[17] Magne Jørgensen and Kjetil Moløkken. How large are software cost overruns? A Review of the 1994 Chaos report. *Information and Software Technology*, 48:297–301, 2006.

[18] Bill Joy and Ken Kennedy. Information technology research: investing in our future. Technical report, President's Information Technology Advisory Committee, February 1999.

[19] Todd Little. Schedule estimation and uncertainty surrounding the cone of uncertainty. *IEEE Software*, 23(3):48–54, 2006.

[20] Michael G. Martin. *Delivering project excellence with the statement of work*. Management concepts, 2002.

[21] Graham Oakes. There's no success like failure. Financial Times, July 12 2006.

[22] Kathleen Ohlson. IT projects are succeeding more often, researcher says. InfoWorld Electric, June 19 1998.

[23] Maija Pesola and Nicholas Timmins. Blackout puts spotlight back on IT failures. Financial Times, November 27 2004.

[24] Joost Schalken. *Empirical investigations of software process improvements*. PhD thesis, VU University Amsterdam, 2007.

[25] Susan Snedaker and Nels Hoenig. *How to cheat at IT project management*. Syngress, 2005.

[26] Theophanis Stratopoulos and Bruce Dehning. Does successful investment in information technology solve the productivity paradox. *Information & Management*, 38:103–117, 2000.

[27] Weidong Xia and Gwanhoo Lee. Grasping the complexity of is development projects. *Communications of the ACM*, 47(5):69–74, 2004.

[28] K.T. Yeo. Critical failure factors in information system projects. *International Journal of Project Management*, 20:241–246, 2002.