SYM-AM-23-106

Excerpt from the Proceedings of the Twentieth Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 10–11, 2023

Published: April 30, 2023

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

Acquisition Research Program
Department of defense management
Naval Postgraduate School

Guiding the Hands of Time: Toward Reliable Schedule Estimates

Charles Pickar—has recently retired from the Naval Postgraduate School (NPS) faculty, where he taught project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is also a retired Army officer with extensive experience in the U.S. defense industry, including Director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. His has also served as Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@gmail.com]

Raymond Franck—retired from the NPS faculty in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included several operational tours, staff positions, and head of the Department of Economics and Geography, United States Air Force Academy. His published work includes many journal articles and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

This paper continues our research agenda concerning advancing the state of the art for estimating defense acquisition program schedules. Accurate schedule estimates provide valuable benchmarks for program managers and reliable dates for the availability of new systems for warfighters. However, most schedule estimates are not correct, with actual time to complete significantly greater initial estimates. This happens for several reasons. One of them is the inherent complexity of modern defense acquisition programs. Another is the generally unfavorable influence of factors outside the program (and program management control). While achieving improved estimates is worthwhile, we also conclude that accurate estimates are generally unobtainable. However, we remain convinced that improvements are possible, which benefit all concerned.

Introduction

Why do weapon system developments always take longer than planned, and why are we always surprised when they do? Why indeed? These two questions have been the core of our schedules research agenda for the past 7 years. Estimating, developing, and executing weapon systems development schedules is rife with challenges, from the schedule estimating process, the complexity of schedules, and the impact of system dynamics on estimation and execution to intangibles associated with our astonishment at the inaccuracies of our schedules. A system development schedule is a promise between the acquisition organization and the customer—and the customer takes us at our word. It is essential to get better.

> Substantial resources have been spent over several decades on improving data and forecasting models. Nevertheless, this has had no effect on the accuracy of forecasts. … This indicates that something other than poor data and models is at play in generating inaccurate forecasts. (Flyvbjerg, 2006, p. 6)

Schedules provide both planned sequencing and duration of weapons system development events. When accurate, the schedule offers the warfighter a reliable date the systems will be available. However, most projects overrun the schedule. This research aims to identify ways to predict the duration of defense acquisition schedules more accurately. This effort is part of a continuing research agenda started by Franck et al. (2016). The latest paper in this line of research (Pickar & Franck, 2022) discussed the issues of commonalities versus individual differences in estimating program schedules, complexities in the processes that determine schedule duration, and program management decisions (which are hard to observe and generally impossible to predict reliably.

It is common knowledge that defense acquisition programs (especially major defense acquisition programs [MDAPs]) experience “deviations” from planned schedules (generally delays). These typically result from untoward events, such as continuing resolution constraints on program progress, or discoveries encountered during program execution. Delays come in many forms, such as deliveries, tests, inadequate system performance, and misspecified tasks (leading perhaps to rework).

However, perhaps the most crucial aspect of this problem is that an adverse development in one aspect of system development can cause undesirable effects on other parts of the program. Thus, for example, the 2004 discovery of weight increases in the F-35 aircraft led to a comprehensive program to pare weight from the aircraft. This included changing methods of assembling the fuselage from snap-together panels to assembly with “traditional fasteners.” This solution to one problem caused other problems, such as manufacturing costs, schedule delays, and lessened F-35 performance (due to losing “good weight”; Pappalardo, 2006).

Also costly (in financial and operational metrics) are delayed retirements of systems that were earmarked to be replaced. As Sweetman (2012) put it regarding the F-35, “The failure of the so-called fifth-generation fighters . . . to arrive on time and cost is having cascading effects throughout U.S. and allied fighter forces.”

Therefore, our overarching research objective is to identify delay-causing events by both examining why the delays happen and proposing tools and processes to better estimate schedule duration at various phases of a given program:

- **Ex ante:** Identify propensity toward untoward developments. Our past work has taken a broad approach to (a) summarize and characterize key elements of the literature on program duration, (b) add to that literature through case study methods, and (c) formulate a more sophisticated model of acquisition program trade-offs. We still think Schedule Estimating Relationships are promising—albeit enriched with modern data science techniques.
- **In medias res:** When the almost-inevitable adverse events occur, it’s helpful to estimate the effects of those events and (ideally) mitigate them. Methods such as computer-enabled content analysis (or text mining) have shown promise in schedule estimation and significant research—including work by acquisition professionals (e.g., Joseph & Sconion, 2020). We also explore prediction markets as another means of crowdsourcing useful information about programs’ progress.
- **Ex post:** What can experience gleaned from past defense acquisition programs enhance our understanding of the basic (albeit complex) processes in play? How might that experience enhance the arts of schedule estimation and program management? We discuss the art of Root Cause Analysis—which the DoD has in place for cost outcomes. We also highlight issues associated with cutting-edge data-gathering and analysis methods.

Our central research question is: *How can the schedule estimating process be improved to reflect the data-identified causes of schedule duration?* Doing that can make program managers more effective. We are also committed to multiple approaches to estimate improvement as appropriate for a significant and challenging problem.

Agenda for Improvement: Ex Ante

Having a realistic initial schedule estimate is vital for many reasons. First is a credible plan for completing the program—within reasonable limits on time and money expended. Such estimates also mitigate management difficulties in programs experiencing challenges.

Second is a reliable plan to bring new capabilities online to operate with other forces. This also facilitates managing the remaining operational life of the "legacy" systems (Sweetman, 2012). This is likely more important in an era of capabilities portfolios (Drabkin, 2019).

Third, a credible schedule provides guardrails for programs encountering difficulties. "How did (program management) know its program (execution) was failing? By the schedule and budget slipping. … If those forecasts were fundamentally unrealistic, a team expected to meet them would fail no matter what they did" (Flyvbjerg & Gardner, 2023, p. 99). Without a reasonable schedule estimate, program management is, in effect, flying blind.

However, there are good reasons why ex ante schedule estimates are commonly not valid (Flyvbjerg, 2006, p. 6) and are much less accurate. First, those best qualified to provide program schedule estimates are, as a rule, optimists who are incentivized to be optimistic (Pickar & Franck, 2022, pp. 12–13).

Second, an acquisition program is a managed effort. Pickar and Franck (2021) discussed the importance of management decisions in determining actual schedule times (pp. 4–12). Since those decisions are unknowable before the actual decision, there's significant uncertainty baked into any schedule.

Third, the difficulty is compounded dramatically by defense acquisition programs' propensity for complexity. Deterministic project scheduling assumes complete information about the scheduling problem to be solved and a static environment within which the schedule will be executed. However, the actual project environment does not behave predictably.

As Dörner (1989) puts it, "Complicated systems . . . derive their complexity from the presence of interrelated variables. One cannot see everything one would like to see" (p. 35). Moreover, complex systems are predisposed to "emergent" behavior (Franck et al., 2012, p. 107; Complex System, n.d.). This implies the system is prone to unpredictable results (Complex System, n.d.)—especially viewed ex ante.

One approach to mitigate this problem is "anchoring" the estimate with reference to past experience: "To create a successful project estimate, you must get the anchor right" (Flyvbjerg & Gardner, 2023, p. 106). Part of "anchoring" is to identify a "reference class," approaching the estimate as "one of a class of similar projects already done" (p. 107). One takeaway from this discussion is that formulating schedule forecasts entails much thought and preparation.[1] Anchoring is a practical step but may not be sufficient for the estimate to be helpful.

One method of anchoring estimates is through Schedule Estimating Relationships, which are generally derived from completed programs. These are (as a rule) ingenious quantitative studies that relate observed program characteristics to actual program schedules.

---

[1] We find nothing inconsistent between the "anchoring" approach and the Scheduling Estimating Relationship methods in acquisition research literature. There is a significant difference in perspective: ex ante versus ex post.

An excellent example of this approach is Light et al. (2018). The authors related actual schedules to various program characteristics, including the acquisition policy era (pp. 3–6). Using those results, a method is available to assess the plausibility of newer schedule estimates based on "program characteristics" (pp. 11–14).

Another is to assume the schedule is not predictable and hedge against that; that is, strive for robust estimates[2] to facilitate resilient program execution. This is no panacea, but as Flyvbjerg and Gardner (2023) put it, those who "lead a big project . . . should . . . protect themselves against overruns. The obvious way to do that is to build a buffer" (p. 9)

These approaches are not mutually exclusive. Current program management practice includes such measures, including contingency and management reserves. Contingency reserves are funds or time set aside to mitigate defined risks. At the same time, management reserves are intended for risks yet to be fully explained (e.g., Project Practical, n.d.), like "black-swan" events (Flyvbjerg & Garnder, 2023, pp. 10–11). A more expansive view would include risks outside the program's scope and, therefore, beyond management control. As noted, original estimates (cost and schedule) are powerfully predisposed to being optimistic. If needed reserves are identified, the overall plan has a degree of resilience that is otherwise missing; management would have the wherewithal to address untoward events.

These should include reserves hedging against developments outside the program. For example, a major acquisition program schedule could consist of a schedule time contingency reserve to account for the effects of restricted funding (and program actions) due to continuing resolutions instead of appropriations.[3]

In Media Res

Since estimating is an inherently uncertain craft, updating program outcomes (especially cost and schedule) is a handy capability. Even more potentially useful is identifying emerging program issues—hopefully in time for program managers to mitigate or forestall them.

"Wise Crowds": Crowdsourcing Acquisition Program Predictions

There is good reason to believe that the collective estimate from a group can be considerably more accurate than the judgment of an expert panel. Substantial experience supports this hypothesis. However, groups can be spectacularly wrong (e.g., financial bubbles, long-shot winners, and black swans). Yet, "even if most people in a group are not especially well-informed or rational, it can still reach a collectively wise decision" (Surowiecki, 2004, pp. xiii–xiv).[4]

A primary framing assumption for groups potentially being intelligent is that every member has private information. Each set of data includes insights and errors (of various kinds).[5] In a "proper" group setting (discussed just below), the (private) errors across the

---

[2] These two approaches are not, of course, mutually exclusive.
[3] One could make a case for including continuing resolutions (CRs) in schedule estimates. The very high probability of a CR in any given could be part of the baseline estimate—with a contingency reserve to address the unknown length of the CRs.
[4] Surowiecki (2004) included an extensive set of notes and citations (pp. 275–296). Due to editorial constraints, we do not delve deeply into that literature here.
[5] One can view each set of private information as having two components: useful knowledge and errors, without individuals being aware of how their private information is divided between those components.

group tend to cancel each other out in the aggregation of opinions, while the private sets of information add to the quality of the collective opinion (Surowiecki, 2004, pp. 10, 41).

Fundamentally, the wisdom-of-crowds hypothesis is a proposition that the capability of the whole is greater than the sum of the individual members' capabilities. Aggregating private knowledge improves the group's capacity to solve problems, while the individuals' errors largely cancel out. For example, Hayek (1945, pp. 17, 19–3) and Smith (1776, pp. 13–16) discussed the ability of a crowd of market participants to reach a sensible economic equilibrium.

Various lines of inquiry have identified characteristics of "wise crowds."

- **Cognitive diversity** is formed in good part by the heterogeneity of private information. Insufficient "cognitive diversity" can lead to "groupthink" and associated pressures to conform (Surowiecki, 2004, pp. 23, 38). Groups that are too much alike find it harder to keep learning because each member is (incrementally) bringing less and less new information to the table (p. 31). Diversity adds different perspectives to the group and lessens the pressure to conform to a consensus, stated or emerging (p. 39).[6]
- **Independence of members** means that individuals are not influenced by other group members. (With insufficient independence, there is a tendency for "herding"). It also promotes a diversity of errors in the sets of private information, which are more likely to cancel out. It also means that each individual's information component is more likely to be additive to the group's information rather than the "same old data" (p. 41).
- **Decentralization** (in an organizational sense) means that information is processed throughout the organization (or outside of it) rather than through a hierarchy.[7] This can foreclose the tendency of hierarchies to filter out information and judgments at lower levels in a structured process to arrive at the "best" answer.

The characteristics above cancel private errors out (or tend, on average, to zero). After the error dross eliminates itself, the valuable sets of personal knowledge make the entire group capable of solving significant problems, such as "cognition" (e.g., election winners), "coordination" (e.g., the operation of a market), and "cooperation" (e.g., getting a disparate group to work together; Surowiecki, 2004, pp. xvii – xviii). This assumes some method of pooling the sets of private information to reach a desired solution. That gets us to the final characteristic.

Aggregation is a process (or set of processes) to bring out a collective assessment related to the entire group's diverse, independent, and decentralized opinions to a good evaluation, forecast, or decision. Aggregating information in a traditional bureaucracy is a well-defined process of screening and assessing information through a series of filters. However, this runs valuable information can be suppressed or disregarded by a hierarchical organization (e.g., the 1986 Challenger launch accident; McCleary, 2023).

Aggregating the collective "wisdom" of a group implies another method other than hierarchical screening, such as prediction markets.

---

[6] For example, the Bay of Pigs Invasion in 1961 was planned and executed "without ever really talking to anyone who was skeptical of the prospects for success" (Surowiecki, 2004, p. 37).

[7] In a hierarchical assessment of available information, perspectives of lower-level individuals generally count for much less than those at higher levels.

Prediction Markets[8]

> The advantage of prediction markets is that they can benefit from the wisdom of crowds. By collecting and weighing the predictions of a large number of traders, they can provide a market-wide forecast that is generally more reliable and balanced than any single expert opinion. (Peters, 2022)

In a prediction market, group members may place bets on defined outcomes. The event may be an election or athletic contest. The group judgment is the prevailing market "price" for the event may be expressed as the probability of a candidate winning or the margin of victory (or defeat). Thus, if Group Member X believes Candidate Y has a 25% chance of winning, he would buy a contract that would be willing to pay up to 25 cents for a contract predicting Candidate Y's victory. The prevailing price is frequently expressed as a point spread in a sports betting market.

An excellent example of a prediction market was discussed in *Nature* (Mann, 2016, pp. 308–310). A "reproducibility project" initiative was formed to determine whether experimental results reported in psychology journals would return the same results with an independent replication. Because the participants "thought it would be fantastic to bet on the outcome(s)," they formed a prediction market to place those bets on whether a series of given results could be replicated. The salient results of this market were that individual experts "hadn't done much better than chance with their individual predictions. But working collectively through the markets, they correctly guessed the outcome 71% of the time" (Mann, 2016, p. 308).

Worth noting is that the group involved were experts in the relevant academic field and could be expected to have significant commonalities in outlook and opinions. Nonetheless, the group outperformed the individuals. Also noteworthy is that a prediction market initiated as an afterthought worked well.

Forms of Prediction Markets

Prediction markets can take many forms by organizing principles and modes of operation. A non-exhaustive list appears below (Peters, 2022).

- **Continuous Double Auction** matches willing buyers and sellers of contracts at a specific market price—much like a stock exchange. The market authority records each transaction.
- **Automated Market Makers** act like a casino or parimutuel betting organization; the "house" serves as the other party to all bets (or trades) and adjusts odds (and payoffs) based on volume for each outcome.
- In **Play Money Markets**, the bets placed convey no market value. Participants are perhaps motivated by reputation or satisfaction in being right.
- A **Decentralized Prediction Market** features trades executed without any central management. "Smart contracts" can then "self-execute . . . to distribute payoffs."

Some Prediction Market Issues Relevant to Defense Acquisition

- **Self-Fulfilling or Self-Negating Group Predictions**: For example, a group prediction of an untoward acquisition program event can lead to management

[8] Due to editorial limits, we provide only an overview of prediction market research results and a sample of the literature. However, the subject is well-recognized as suitable for in-depth research. See, for example, *The Journal of Prediction Markets* at https://www.scienceopen.com/collection/755392c6-34de-437c-9d0a-c768f6f128bb

actions to prevent that event. We discussed this issue in a previous paper (Pickar & Franck, 2022, p. 24).

- **Positive and Perverse Incentives**: Prediction markets look like and can operate like betting markets. As such, there can be incentives to engineer a favorable outcome, which has happened in sports betting operations. This problem can be addressed by limiting the stakes. For example, the Iowa Electronic Market limits positions to $500 (University of Iowa, n.d.). The Reproducibility Project gave each group member $100 to wager (Mann, 2016, p. 308). Hence, relatively small stakes can nonetheless elicit candid assessments.[9] And that is good news for acquisition prediction markets.

However, motivated participation is beneficial. As one observer put it, "I can create a poll that can mimic everything about a prediction market, except markets, have a way of incentivizing you to come back at 2 a.m. and update your answer" (Mann, 2016, p. 310). Most importantly, a well-functioning prediction market provides valuable incentives for participants to reveal their judgments regarding the event at hand sincerely.

Prediction Market Issues

As good as prediction markets can be, there are some problems. Typically, they focus on well-defined binary outcomes (e.g., win or lose) that occur at a definite time. Results of sporting events and elections are good examples of this category.

But what happens if the outcomes are more complicated? Suppose a wise group identifies an emerging problem in an acquisition program (such as a schedule slip). Suppose also that alerted program management undertakes a remedy that averts the crisis.[10] How do the prediction market rules determine the winner?

One way around this problem is to have more detailed results. For example, group members could choose an outcome in perhaps two parts. Will the group identify the particular problem? If so, will management action avert the problem? While this seems a reasonable solution, even more, complex bets might arise in a well-designed prediction market for an acquisition program.

There might also be dilemmas (or trade-offs) in prediction market design. An essential assumption for prediction markets harnessing the wisdom of crowds is the "marginal trader" (Adam, 2016, p. 310), one who acts to benefit from current group misconceptions.[11] Incentivizing effective marginal traders might entail substantial incentives to be correct. Doing that could, in turn, constitute a significant incentive to take action (unethical or illegal) to increase the odds of winning the bet.

Another obvious issue is that acquisition programs (especially MDAPs) are lengthy and have uncertain termination or milestone dates. Defense acquisition prediction markets operating arena will likely need special care in framing the questions upon which to place bets.

---

[9] There's reason to believe that simply being proved right is a useful incentive for thoughtful participation. As Kathryn Schulz (2010) put it, "The experience of being right is . . . one life's cheapest and keenest satisfactions" (p. 4).

[10] This is not so far-fetched. Miller (2012) offered a method that can yield actionable indication of an emerging problem in acquisition programs (pp. 48–49). It's therefore reasonable to suppose that prediction markets might also provide similar warnings.

[11] Beaton and Cohen (2023) offered advice for would-be marginal traders in this year's NCAA Men's Basketball Tournament, noting that "there's value in finding the best, least popular team and making it your national champion." However, their championship pick, UCLA, lost in the Round of 16.

The issues and problems we've raised are untested hypotheses but could add to the practical difficulties of organizing a functional prediction market. As such, they appear to be matters for more research and experience.

Acquisition Data Qualitative Analysis

As part of our ongoing study this year, we apply qualitative research methods to improve on a 2018 macro-level study of factors that define schedule delays (Pickar, 2018). The 2018 analysis of Selected Acquisition Reports (SARs) used a cumbersome, manual process to code each schedule explanation text entry to convert it to structured, measurable data.

A constraint of studying the schedule process is data availability and data analysis techniques. Data for this project come from the DoD SAR. The documents to be examined are the SAR sections on the executive summary and the schedule change explanation of the SARs at the Washington Headquarters Services website (DoD, n.d.). The SARs are reports to Congress and, as such, are text or unstructured data. Text mining and Computer Assisted Qualitative Data Analysis Software (CAQDAS) are two ways to analyze unstructured data. CAQDAS allows qualitative analysis of hundreds to thousands of documents. Text mining, on the other hand, offers analysis for millions of documents. For this paper, we consider CAQDAS and text mining to be interchangeable. We are attempting

> the discovery by computer of new, previously unknown information by automatically extracting information from different written resources. A key element is the linking together of the extracted information together to form new facts or new hypotheses to be explored further by more conventional means of experimentation. (Hearst, 2003, p. 1)

Qualitative Analysis Method

This qualitative analysis uses a modified version of the "grounded theory" method. Grounded theory is "the discovery of theory from data systematically obtained and analyzed in social research" (Glaser & Strauss, 2017, p. 2). The grounded theory approach uses an iterative data collection and analysis process as central to theory development. In this case, studying the events described in the SARs could contribute to a model of schedule activities in ACAT I programs examining causes for both delay and acceleration. Schedule delays often consist of more than one root cause, and the dynamics of the delays frequently cause further difficulties in schedule and elsewhere. Grounded theory is practical when no existing theory explains the activities being observed. The process steps include collecting the data; analyzing the data; identifying/grouping the discovered concepts; and finally, identifying any relationships between categories of data.

A necessary pre-analysis step required collecting and preparing the data. Preparing the data involved resolving/removing duplicate data, fixing structural errors, and identifying outliers. Once accomplished, the qualitative analysis started with a literature review of the broad area of scheduling.

Qualitative Analysis Literature Review

We used a qualitative, two-part process to examine the literature on schedule. The first part was a broad examination of 83 peer-reviewed papers on schedule. The literature review identified themes for use in the CAQDAS analysis for coding. Figure 1 is a treemap of the initial analysis of those papers. Of the top five themes, uncertainty (a subject we have not examined) ranked fifth in the literature. In this literature review, uncertainty included both schedule risk and schedule uncertainty. The literature ranged from uncertainty impacting the schedule development process from estimates to management to execution. In some papers, the terms are used interchangeably. We then conducted a separate literature review on uncertainty.

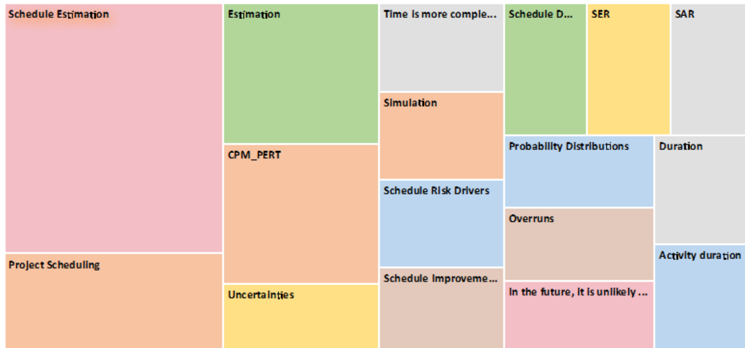


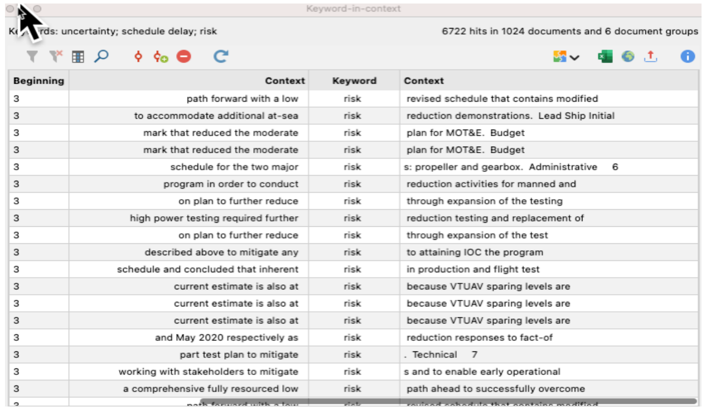
Figure 1. Treemap of the Schedule Literature

This review also provided the initial word cloud, a visual representation that allows the organization of the unstructured data to classify themes and find the relationships between those themes. Figure 2 is the word cloud for the uncertainty literature review.

Figure 2. Word Cloud for Uncertainty Literature Review

Complexity was also identified in this literature review in the context of uncertainty. The uncertainty literature review identified suitable variables/codes for further analysis. While many scholars equate complexity with uncertainty (e.g., Williams, 2002), we separate them.

Risk is also a frequent context word often used interchangeably with uncertainty. DoD project managers (PMs) are well-versed in risk management and very familiar with uncertainty. Some argue the essence of project management is actually about managing risk. Risk, its management, and in some cases, avoidance is central to good project management. Uncertainty, however, is a less appreciated but significant factor in DoD project management. Arguably, it can cause more headaches. Figure 3 shows the risk keyword in context. This CAQDAS tool provides a way to understand how the term risk is used and whether the authors are substituting (in this case) risk for uncertainty.



Beginning	Context	Keyword	Context
3	path forward with a low	risk	revised schedule that contains modified
3	to accommodate additional at-sea	risk	reduction demonstrations. Lead Ship Initial
3	mark that reduced the moderate	risk	plan for MOT&E. Budget
3	mark that reduced the moderate	risk	plan for MOT&E. Budget
3	schedule for the two major	risk	s: propeller and gearbox. Administrative 6
3	program in order to conduct	risk	reduction activities for manned and
3	on plan to further reduce	risk	through expansion of the testing
3	high power testing required further	risk	reduction testing and replacement of
3	on plan to further reduce	risk	through expansion of the test
3	described above to mitigate any	risk	to attaining IOC the program
3	schedule and concluded that inherent	risk	in production and flight test
3	current estimate is also at	risk	because VTUAV sparing levels are
3	current estimate is also at	risk	because VTUAV sparing levels are
3	current estimate is also at	risk	because VTUAV sparing levels are
3	and May 2020 respectively as	risk	reduction responses to fact-of
3	part test plan to mitigate	risk	. Technical 7
3	working with stakeholders to mitigate	risk	s and to enable early operational
3	a comprehensive fully resourced low	risk	path ahead to successfully overcome

Figure 3. Risk Keyword in Context

The eminent economist Frank Knight (1921) proposed one of the classic distinctions between uncertainty and risk by arguing that risk can be assigned probabilities, making it theoretically predictable. Still, uncertainty is unpredictable (Knight, 1921). PMs (both DoD and industry) seldom have complete information, but current scheduling techniques assume full knowledge. This includes assumptions that the development environment is predictable. Uncertainty begins in the scheduling process with estimates of task duration. Those estimates are usually rolled into the schedule development process, a software package built on the critical path method (CPM) and the Program Evaluation Review Technique (PERT). For the PM, a developed schedule seems to embody predictability. Of course, nothing is predictable in the real world of system development (Demeulemeester & Herroelen, 2006; Dörner, 1997). In fact, "The more we know, the more clearly we realize what we don't know" (Dörner, 1997, p. 93).

Project management uncertainty is "the inability to predict future outcomes" (Shenhar & Dvir, 1996, p. 610). PMs crave certainty; however, the opposite is ever present in weapons system development projects. In fact, uncertainty in project management is generally assumed in the development of high-tech weapons. Uncertainty manifests in many other aspects of project management, including uncertainty in funding and the workforce (both skills and labor issues), to name a few.

On the other hand, risk is not only managed; it is quantified with probability (Koleczko, 2012). Uncertainty can't be quantified. Risk is a distinct and identifiable project influence (and a practice embedded in the project management process.)

Like risk management, dealing with uncertainty depends on the program manager's tolerance for ambiguity (Koleczko, 2012). Further, decision-making relates to certainty, risk, and uncertainty (Kerzner, 2013). Finally, too much information creates uncertainty, an idea all too familiar for combat commanders and staff. Uncertainty drives the need to gather more information, and that information inhibits decision-making in a vicious cycle—of paralysis by analysis.

Coding & Variables

The last part of this qualitative analysis applies the codes extracted from the literature review to facilitate sensemaking. Some liken qualitative research to finding the "needle in a haystack," an apt analogy. The SAR schedule database alone contains more than 3,000 records with over 585,000 words—an impossible task if one sets out to read, comprehend, and analyze without computer assistance.

Figure 4. Word Cloud for SAR Schedule

Qualitative Methods Summary

This study used a qualitative analysis approach to automate the analysis of the SARs. A qualitative analysis approach potentially improves our understanding of schedule delay in ways beneficial to program managers. Part of this study used CAQDAS to perform a broad literature review of system development schedules and a more focused examination of schedule uncertainty. Both permitted us to refine the results of our 2018 research effort.

- The automation provided by the CAQDAS software significantly improves the ability to search and comprehend large amounts of unstructured data (both literature reviews and SARs data). We noticed improvements in the time necessary to do the research and the accuracy provided by the abovementioned tools. In fact, the software uncovered minor errors in the original manual study that had remained unnoticed.
- One of the difficulties in the original study was understanding the differences in terminology used by the different authors/program offices. For instance, some authors would say, "The schedule was updated to reflect actual dates." Others did not use that phrase but referred to the same idea of modifying the report to reflect the actual date something occurred. Still, others simply noted an update to the scheduled dates. The software provides the ability to identify these similar types of activities. Still, since they can be examined side-by-side, it gives the ability to be more accurate in assessment.
- Using uncertainty as the central theme for coding the documents provided insight into the different perceptions of program offices between uncertainty and risk. Uncertainty, as a code, occurred 27 times in the executive summary of the SAR and 26 times in the schedule change exclamation part of the SAR. A concern in the original manual analysis was the duplication of different terms. In other words, CAQDAS provided the ability to identify the existing programs and issues being discussed in the schedule context. This provided a check on the accuracy of the counts.
- The further coding of uncertainty required an understanding of the different authors' uses of the word uncertainty and synonyms for uncertainty. For example, some authors describe situations as ambiguous in the context referring to uncertainty. As we continue our analysis across the complete SAR, understanding the actual terms being used will become one of the critical activities.
- Examination of the concept of uncertainty provided another further classification of schedule delay. We created a category to track whether the source of the delay was internal (by the PMO) or external (outside the PMO). Thirteen percent of the data records were classified as having causes external to the PMO. Examples of external

change reasons included contract negotiation delays; contract completion delays; follow-on contract award delay; testing delays based on the availability of the testing unit; service changes in start dates; delays in contractor delivery, installation, and completion; and labor activities (strikes, etc.).

- This approach provides granularity in the administrative factors category. In the initial analysis, administrative reasons were more than 30% of the entire analysis. Administrative includes updates to APB, ADM changes, and changes resulting from Nunn–McCurdy processes, program restructuring, and a general category.

Ex Post Analysis

Those who execute acquisition programs seldom clearly understand the situation's dynamics. As Dörner (1989) put it, "A tendency, under time pressure, to apply doses of established measures. . . . An inability to . . . properly assess the side repercussions of one's behavior" (p. 33), and "a tendency to think in terms of isolated cause-and-effect relationships" (p. 35).

Root Cause Analysis (RCA) is an ex post method intended to provide a detailed understanding of the process that led to a bad (or excellent) outcome. The intention is to discover how to remedy (or replicate) similar processes. RCAs set out to find the complete chain of events that led to the result in question (good or bad).

RCA is practiced in many fields, such as manufacturing process control, medical procedures, and accident investigations. It consists of answers to three questions.

- **What happened?** While the result is generally self-evident, the sequence of events that produced the result typically involves careful study.
- **Why did it happen?** What set of causal relationships ties those events together? The primary intent is to discover the chain of events to identify proximate and root (or underlying) causes.
- **What should we do about it, or what can we learn from it?** These "takeaways" involve measures to prevent (replicate) similar occurrences of bad (good) results.

RCA of cost growth has also been practiced in defense acquisition programs, with a lead organization (e.g., OSD, PARCA) mandated by legislation.[12] Among its commissioned cost RCAs were studies of MDAPs undertaken by Institute for Defemse Analyses (Arnold et al., 2010) and RAND (Blickstein et al., 2011).

One example from commercial aerospace is the fate of the Boeing 737 MAX airliner program. Program difficulties became evident with two catastrophic crashes (*what happened*)—in October 2018 and March 2019. Pickar and Franck (2020) took an acquisition-schedule perspective in their RCA.

Why it happened (from a development schedule perspective) is summarized in Figure 5. The decision (about 2010) to delay the Boeing 737 replacement program until 2020 allowed Airbus to steal a march in the form of a re-engined 320 family. The rapid commercial success of the new variant (A320neo) forced Boeing management to promise a new (more efficient) variant of the 737—promised for delivery at a highly optimistic date. Adapting the Boeing 737 airframe to accommodate new and larger turbofans on a tight schedule included (a) moving the engines forward and higher and (b) resolving the attendant aerodynamic complications with software fixes rather than airframe modifications.

---

[12] This function is now within the Office of Acquisition Data and Analytics.

Pickar and Franck (2020) traced the causal chain back to Boeing's decision to delay the Boeing 737 replacement program: "The fundamental root cause of the current 737 MAX difficulties was a strategic miscalculation a decade in the past" (p. 11). Boeing had a severe problem in two parts: "(1) coaxing additional competitive life from a half-century-old design and (2) doing so in a manner responsive to the threat posed by the Airbus A320neo" (p. 11). In short, the 737 MAX program was begun "late to need." Boeing could have avoided this problem by starting a 737 replacement program around 2010.

What could also be learned was the dangers of rigidly specifying total program times (Pickar & Franck, 2020, pp. 11–12): "aspirational" schedules, which "are driven by political and commercial processes and decisions."[13] It is an example of making engineering development fit a strategy, rather than allowing the engineering (tasks) to define the time needed.

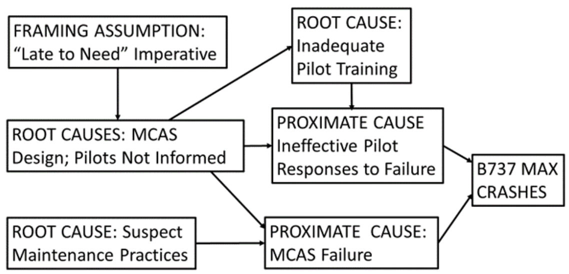


Figure 5. A Root Cause Analysis of the Boeing 737 MAX Accidents of 2018, 2019 (Pickar & Franck, 2020, p. 9)

Concluding Comment

We believe current methods (even improved) acquisition schedule estimates are unlikely to solve this thorny problem completely.

We've discussed several reasons for this, including those summarized below.

- Those most knowledgeable about the program at the start are likely those most optimistic about the program. Furthermore, competitive bidding processes tend to reward that innate optimism. This can also result in "winner's curse" outcomes in which the chosen supplier can be the most optimistic—and, therefore, the most likely to be inaccurate.
- MDAPs are systems with many moving parts that interact with each other in ways not known in advance and imperfectly known during program execution. This complexity seems inherent in defense acquisition programs.
- Program execution is a managed process in which management is expected to make decisions with trades (perhaps implicit) among schedule, cost, and performance. Each decision is generally impossible to predict and usually tricky to discern ex post.

A modest proposal: Given the innate degree of uncertainty in any program schedule, a helpful benchmark for a schedule estimate might be the answer to two questions. First, what's the schedule duration if everything goes right?[14] Second, how much do we wish to hedge that bet (through cost and schedule "reserves")?[15]

[13] Acquisition programs starting "late to need" is not a rarity. The new US ICBM (GBSD) and Airborne Early Warning Aircraft (E-7) are contemporary examples.
[14] The DoD tends to get these anyway.
[15] We view this as more "truth in advertising" than a new approach.

References

Arnold, S. A., et al. (2010). *WSARA 2009: Joint Strike Fighter root cause analysis* (IDA Paper P-4612).

Beaton, A., & Cohen, B. (2023, March 14). Want to win your NCAA tournament bracket pool? Pick this team. *Wall Street Journal*. https://www.wsj.com/articles/march-madness-bracket-predictions-ucla-464d0524

Blickstein, I., et al. (2011). *Root cause analyses of Nunn-McCurdy breaches Volume I: Zumwalt-Class Destroyer, Joint Strike Fighter, Longbow Apache, and Wideband Global Satellite*. RAND Corporation.

Complex system. (n.d.). In *Wikipedia*. Retrieved February 9, 2023, from https://en.wikipedia.org/wiki/Complex_system

Demeulemeester, E. L., & Herroelen, W. S. (2006). *Project scheduling*. Springer Science & Business Media.

DoD. (n.d.) *Washington Headquarters Services*. Retrieved April 6, 2023, from https://www.esd.whs.mil/FOIA/Reading-Room/Reading-Room-List_2/Selected_Acquisition_Reports/

Dörner, D. (1989). *The logic of failure* Translated by R. & R. Kimber, Pegasus Books.

Dörner, D. (1997). *The logic of failure*. Basic Books.

Drabkin, D. (2019). Section 809 Panel recommends reforming defense acquisition with updated structures, simplified procedures and an empowered workforce. *DAU News*, 31.

Flyvbjerg, B. (2006). From Nobel Prize to project management: Getting risks right. *Project Management Journal*, *37*(3), 5–15.

Flyvbjerg, B., & Gardner, D. (2023). *How big things get done*. Currency.

Glaser, B. G., & Strauss, A. L. (2017). *Discovery of grounded theory: Strategies for qualitative research*. Routledge.

Hayek, F.A. (1945). American economic review. In H. Townsend (Ed.), *Price theory: Penguin modern economics readings* (pp. 17–31).

Hearst, M. (2003). *What is text mining?* University of California, Berkeley.

Kerzner, H. R. (2013). *Project management*. John Wiley.

Klein, G. (2004). *The power of intuition: How to use your gut feelings to make better decisions at work*. Currency. https://books.google.com/books?hl=en&lr=&id=okyNDQAAQBAJ&oi=fnd&pg=PP15&dq=the+power+of+intuition&ots=eRTyAR5S3r&sig=-xqpMc0W6CI7YYG4fYxW-a-ltxI

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Houghton Mifflin Company. https://cir.nii.ac.jp/crid/1573387450829321344

Koleczko, K. (2012). Risk and uncertainty in project management decision-making. *Public Infrastructure Bulletin*, *1*(8), 1–8. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fbbe5a5a8da599f31b641fa78ca9a3bda44817dd

Light, T., et al. (2018). *Benchmarking schedules for major defense acquisition programs* (Document No. RR-2144-AF). RAND. https://doi.org/10.7249/RR2144

Loxton, M. H. (2021). *Using MAXQDA for strategic foresight*. MAXQDA Press.

Mann, A. (2016, October 20). Market forecasts: Prediction markets can be uncannily accurate. *Nature*, 308–310. https://www.doi.org/1038/538308a.PMID277622382

McCleary, R. M. (2023, January 27). The man who tried to stop the Space Shuttle Challenger's launch. *Wall Street Journal*. https://www.wsj.com/articles/the-man-who-tried-to-stop-the-challenger-launch-space-shuttle-exploration-roger-boisjoly-moral-injury-11674857494

Miller, T. P. (2012). *Acquisition program problem detection using text mining methods* [Thesis, Air Force Institute of Technology].

Pickar, C. K. (2018). *Informing DoD program planning through the examination of the causes of delays in acquisition using acquisition data*. https://calhoun.nps.edu/bitstream/handle/10945/58762/SYM-AM-18-079-019_Pickar.pdf?sequence=1

Pickar, C., & Franck, R. (2020). *It's about time: Toward realistic acquisition schedule estimates* (SYM-AM-20-051). *Proceedings of the Seventeenth Annual Acquisition Research Symposium*. https://event.nps.edu/conf/app/researchsymposium/unsecure/diprop646/52

Pickar, C., & Franck, R. (2021). *Telling time: Getting relevant data for acquisition schedule estimating relationships* (NPS-AM-22-013). Naval Postgraduate School.

Pickar, C., & Franck, R. (2022). *Time for a change: Data-driven schedule relationships* [Paper presentation]. Western Economic Association Annual Conference.

Project Practical. (n.d.). *Contingency reserve vs management reserve: PMP concept*. Retrieved March 1, 2023, from https://www.projectpractical.com/contingency-reserve-vs-management-reserve/

Schulz, K. (2010). *Being wrong: Adventures in the margin of error*. HarperCollins.

Smith, A. (1776). An inquiry into the nature and causes of the wealth of nations. In E. Canaan (Ed.), *Modern library*.

Surowiecki, J. (2004). *The wisdom of crowds*. Doubleday.

Sweetman, B. (2012, November 26). Generation gap. *Aviation Week*, 22–23.

Torenberg, T. (2021). *A primer on prediction markets: Ideas and musings*. https://eriktorenberg.substack.com/p/a-primer-on-prediction-markets

University of Iowa. (n.d.). *What is the IEM?* Retrieved January 23, 2023, from https://iemweb.biz.uiowa.edu/media/summary.html

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net