A Real World Calibration Interval Analysis Program

Speaker/Author: James M. Ingram Jr.
Metrology Management Consultant
505 Wadhams Street
Ogdensburg, New York 13669
(315) 393-0229
jingram@twcny.rr.com

Abstract

There are in excess of 28,000 pieces of TMDE owned by the Royal Saudi Air Force (RSAF), consisting of over 7000 different model numbers requiring calibration. While the calibration intervals as used by the US Air Force (USAF) were adequate for over twenty years, recent calculations of reliability indicate lower than desired numbers. A new system was developed for determining calibration intervals specifically for the conditions of use and maintenance found in Saudi Arabia. With limited budgetary, manpower and computer resources available, a creative yet valid method had to be set up to return the pool of Test, Measurement and Diagnostic Equipment TMDE to the desired level of reliability. This paper describes the development and implementation of that system.

Introduction

The RSAF established their metrology program with the assistance of and based on that of the USAF. This was a less expensive way to start without having to reinvent all of the processes. This emulation included management processes, organizational structure, selection of laboratory standards and the use of USAF calibration and management procedures. Subscribing to USAF procedures established the same conditions for calibration of the TMDE and would allow the use of calibration intervals as found in USAF indexes. The foundation for the USAF intervals is a database of maintenance information collected from the USAF worldwide metrology organization.

This situation may have continued except for a series of analysis done by the Metrology Engineering group using maintenance data on the RSAF TMDE collected over the years. In December of 1999, a first review was made to determine if the 85% reliability target was being met. At that time, it was obvious that there were some serious shortfalls in reliability. For some model numbers the observed reliability was found to be as low as 15 or 20%, and isolated models even lower. This was not acceptable and something had to be done to improve the situation. The latest RSAF Metrology Management instructions stated that the intervals would be adjusted based on the observed reliability found on TMDE used in Kingdom. This procedure had never been implemented, and all intervals were still being determined with USAF and Royal Air Force (RAF in the UK) information. In order to implement the use of the RSAF determined calibration intervals it was necessary to obtain written approval from RSAF headquarters. A second review was made in the summer of 2000 and used as justification to RSAF Headquarters for permission to start making changes. This permission process took almost a year.

The Playing Field

One of the major difficulties with determining calibration intervals for RSAF TMDE is the proliferation of so many different model numbers. A review of the inventory determined that there were over 28,000 pieces of TMDE. Unfortunately there were well over 7,000 different model numbers, each with its set of history records. A random sample of 300 items determined that the weighted average interval was 14 months, and the average quantity of any model number was only four. The RSAF instructions required that there be 30 calibrations prior to making an interval change. From the information found in the inventory survey it was obvious that it would take an average of over 8 years to obtain sufficient data to make the first change, then, of course, start counting again to make another. This was not practical and would not really solve the reliability problem. A method had to be established to enable the evaluation and adjustment of calibration intervals with much smaller numbers of calibration actions.

The central computer system in use provided only three useable reports for making calibration interval studies. Two reports provided total calibration actions, stated how many were received in tolerance and how many needed repair or adjustment. These reports included the currently assigned calibration interval and its source, whether USAF or RAF. In reality, the third report was a full inventory printout showing all instruments in the inventory. This report made it possible to determine how many of any particular model number were in the TMDE pool.

Queries to the computer department determined that any modifications made to the interval change algorithm in use would take approximately five years. In fact, to design a new report was considered a change to the system and required several years to accomplish. We were told to live with what we had. These facts determined how we would proceed. It was obvious the method needed would have to depend on the available data.

The only computers available were the personal computers of the engineers in the group. There was no budget to procure special calibration interval analysis software. We were on our own. This meant that the Metrology Engineering group had to complete the evaluation process for over 7,000 models, along with their normal responsibilities, with minimal tools. It was not possible to stretch the process out over many months. While these engineers dealt with calibration intervals as a matter of course in their jobs, only one engineer had any in-depth knowledge of the calibration interval adjustment process. We required a process that would be easy to follow. Several training sessions would also be necessary to establish a base level of knowledge of the statistics involved as well as the assumptions we would make in the interval analysis project.

Building a Process

The available data gave us a binary value for the calibration results, the as-found condition was either in-tolerance or it required adjustment or repair. We did not know a time frame for this data only the total number of failures and a value for Observed Reliability, R_O (percent found in-tolerance) since the beginning of data collection. We had various size populations of instruments with long and short calibration intervals generating various amounts of data. Any review process must handle all of these situations. After review of the requirements provided by NCSL in Recommended Practice RP-1 [1], it was determined that we would need to use Method A3. Unfortunately, we were missing the time and resources to do the lengthy math required on each of over 7000 model populations. Two questions needed to be answered: Does the model population need an interval change? What is the appropriate new interval?

Provided with RP-1 is the program, *Interval Test*, for evaluation of calibration intervals. This program allows the user to take the number of calibrations and the number found in-tolerance to determine if the current calibration interval is acceptable. *Interval Test* is designed to work well with small populations of data, and thus fits our situation where over 4000 of the model numbers had less than 10 calibrations. Only 15% had at least 30 calibrations. Given this tool, the next step was to develop a set of criteria to answer the first question in our project, “Does the model population need an interval change?”

Several requirements were given as goals of the RSAF Metrology program. The Target Reliability (R_T) was to be 85%. The first decision was made to determine the level of confidence we needed for this evaluation. *Interval Test* bases its decision on Upper and Lower Confidence Limits determined statistically from the pass-fail data. The decision was made to use 95% confidence limits. Figure 1 shows a sample *Interval Test* data screen.

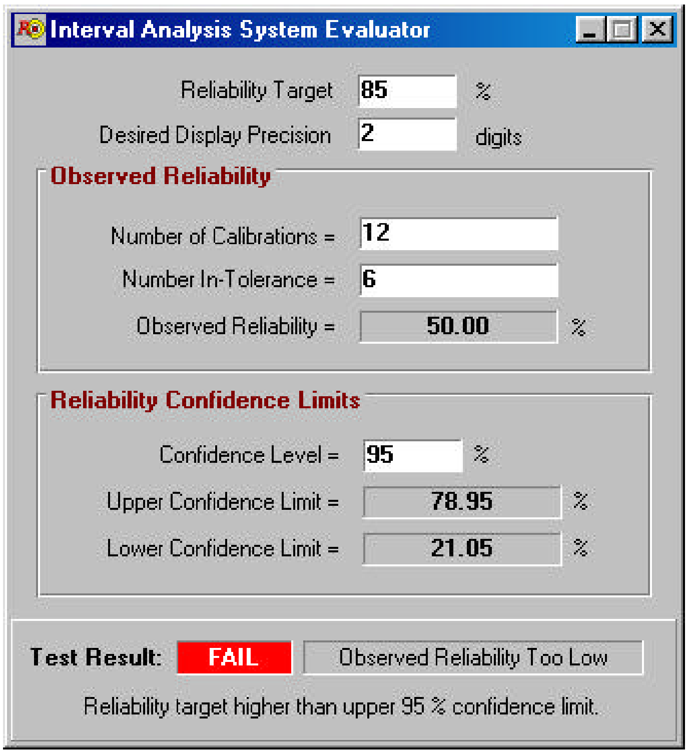

Figure 1. RP-1, *Interval Test* data screen.

This example shows the entry of R_T as 85% and our required confidence level as 95%. This example shows 12 calibrations with only 6 instruments found to be in-tolerance. The program calculates the values for R_O and both the UCL and LCL. The result indicates that our current calibration interval fails because the R_O is too low.

The next step in developing our process was to use *Interval Test* to establish thresholds of pass/fail for the various sizes of data populations. A sensitivity analysis was performed using *Interval Test* for all data population sizes from 7 to 29 to determine the minimum number of in-tolerance calibrations necessary to pass the test.

The results of this analysis were placed in a process flow chart or decision tree that would allow the engineer performing the evaluation to quickly determine if a change to the interval would be required. Figure 2 on the next page shows a similar chart.

Studying the chart it is obvious that there are also some blocks in addition to those devoted to the use of *Interval Test* to determine a pass/fail for the currently assigned interval. These blocks involve answering the question, “What is the appropriate new interval?” Several decisions were made regarding when and how to change the calibration interval. The first decision was to attempt to fine-tune those cases with over thirty calibrations. We chose to make a change whenever the R_O was not within the center 50% of the confidence interval for a 95% confidence level. We could have just reduced the target confidence interval but chose to remain at the 95% level to maintain consistency and reduce human error in switching between cases. The next decision was to not make a change in the very small data sets of 9 calibrations or less, unless we had only one or two instruments of the particular model number. This required that each had at least 3 calibrations. We were instructed to do nothing in cases where we had 6 or fewer calibrations.

For cases where an increase in interval was indicated we established two criteria. The first was to make changes for the large data cases, thirty or more calibrations, by tweaking the interval to within the central 50% of the confidence interval. The second was that we only consider an increase for those cases with 20 to 29 calibrations where the R_O was $\geq$ 95%. We made no increase in cases with less than 20 calibrations.

In each case of a change, an analysis of the maintenance records was made to determine any maintenance problems as well as the advisability of making a change. Thus this area became subjective with the goal of improving reliability.

The RSAF reports, providing the calibration data, also provided the recommended calibration interval as determined by the embedded algorithm. The mechanism was designed for cases where there were 30 or more calibrations but gives a recommendation regardless of the number of calibrations. The simple formula allowed a percent change to the interval based on how close the value of R_O was to the R_T., i.e. if the R_O = 65% and R_T = 85%, the report advised a 20% change in the interval. This would reduce an interval of 9 months by 1.8 months to a new interval of 7 months. We used these suggested intervals as a range of possible changes for the case on hand. It was then left up to the engineer to make a judgment call based on experience with the actual instruments and all relevant maintenance data and calculations to decide how much to change the interval. Training sessions were held to review various cases to establish a decision process as consistent as possible.

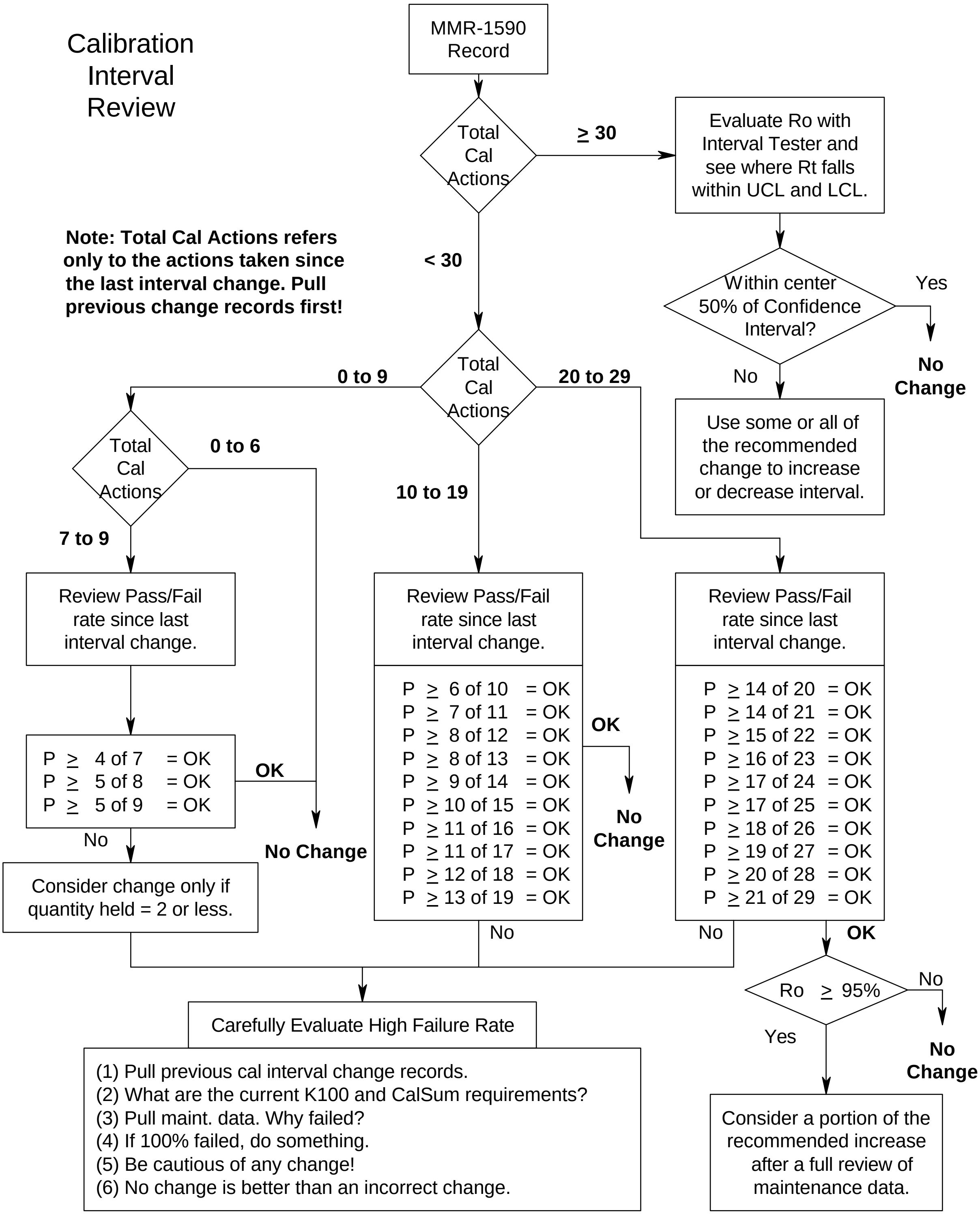

Figure 2. Calibration Interval Review Decision Process

Execution

Once the process development and engineering training were complete, it was time to execute. Fresh data reports were generated in May 2001. The review process for the entire inventory took approximately 80 man-days. It resulted in just over 400 required changes. A paper record was made of each change required that included a picture of the data at the time of evaluation. This was necessary, as the maintenance data collection system did not allow for restarting the data counting. In other words, we needed this information during the next review; the total calibrations reported would always start from the beginning of the maintenance history. As a check for consistency in decision-making the chief engineer then reevaluated all targeted changes. The interval changes were then manually entered into the RSAF maintenance data system.

This process was repeated eight months later in January 2002. This resulted in approximately 200 more changes with only one case requiring a second change to an even lower calibration interval.

Conclusions

While it is too soon to tell if the changes have made a difference, it was a necessary exercise since many of the pieces that required a changed interval might have an effect on aircraft maintenance operations. Incorrect readings and setup during maintenance could possibly cost lives in the case of a catastrophic aircraft failure.

The system, though crude, allowed the necessary changes to be implemented. Since the inception of this process a new version of *Interval Test* has been developed that makes the rigorous calculations necessary to properly determine the recommended calibration interval. *Method A3 Interval Tester* [2] implements the full requirements from RP-1. The additional information input to the program on the current interval in use allows for the calculation of a new interval.

Where do we go from here?

Looking to the future we can apply some of the lessons learned and make changes designed to improve our available data. The following suggestions would be a start.

1. Change to the new program, *Method A3 Interval Tester*, is necessary before the next round of reviews. A new decision chart should be prepared for the use of this new software.

2. Clean up the database and consolidate the model numbers where possible to increase some of the populations. Some are just the result of different options having no relationship to the reliability or calibration of the instrument. Many valid model numbers are the same number written in a different format.

3. Remove obsolete numbers from the database. Currently, along with the 7000-8000 valid model numbers, there are over three thousand obsolete numbers. These cause a lot of

unnecessary work, as the reports do not indicate any obsolescence and the extra 3000 are evaluated along with the rest

4. Provide training to the technicians generating the data to establish consistent as-found condition coding.

5. Start the long-term process of making changes to the central maintenance data collection program to allow for the use of the A3 Method as well as new reports providing better monitoring of the TMDE reliability.

References

1. *Recommended Practice RP-1, Establishment and Adjustment of Calibration Intervals*, Third Edition, National Conference of Standards Laboratories, January 1996. Available from NCSL International, 1800 30th Street, Suite 305B, Boulder, CO 80301.

2. *Method A3 Interval Tester*, Integrated Sciences Group, April 2001. Available as Freeware from the ISG website at http://www.isgmax.com/.