Top 10 Spreadsheet Risks
Excel Community
Blog series summary
It is estimated that over 90% of spreadsheets contain errors (see section 9).

This document, which is a summary of blog postings from the Excel Community, outlines the top 10 spreadsheet risks – how to identify them and how to avoid them. Half of the risks are due to human factors and half are due to technical engineering issues.

The blog series has been written by Excel specialist Grenville Croll, past chair of EuSpRIG (the European Spreadsheet Risks Interest Group).

The top 10 spreadsheet risks are as follows:

<table>
<thead>
<tr>
<th>Risk</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overconfidence</td>
<td>01</td>
</tr>
<tr>
<td>Failure to specify</td>
<td>02</td>
</tr>
<tr>
<td>Ignorance</td>
<td>03</td>
</tr>
<tr>
<td>Failure to test</td>
<td>04</td>
</tr>
<tr>
<td>Denial</td>
<td>05</td>
</tr>
<tr>
<td>Failure to control</td>
<td>06</td>
</tr>
<tr>
<td>Arrogance</td>
<td>07</td>
</tr>
<tr>
<td>Failure to document</td>
<td>08</td>
</tr>
<tr>
<td>Failure to review</td>
<td>09</td>
</tr>
<tr>
<td>Inefficiency</td>
<td>10</td>
</tr>
</tbody>
</table>
Almost exactly 10 years ago, Professor Ray Panko from the College of Business Administration at the University of Hawaii, flew half way round the world to deliver a truly memorable paper about Reducing Overconfidence in Spreadsheet Development at the fourth EuSpRIG conference in Dublin, Ireland.

There is a broad literature on overconfidence, which occurs in many areas of human life. When driving motor vehicles, one of the main causes of accidents is because overconfident drivers do not take avoidance measures until it is too late. Surprisingly, engaging in risky behaviour in motor vehicles can be self-reinforcing. If we take risky actions when we drive this only rarely causes accidents. So we don’t change our risky behaviour because we lack the evidence that such behaviour is indeed risky. We are not in possession of the full facts of our aggregate behaviour as our insurers most certainly are, and so we continue our less than optimal behaviour.

This is similar with spreadsheet development. Because so few people formally test their spreadsheets properly, they don’t find any or many errors. This encourages them to perpetuate their behaviour. Because we don’t look carefully to find mistakes in our spreadsheet endeavours we become ever more confident that our spreadsheets are perfect despite the overwhelming evidence against.

Ray Panko systematically tested this overconfident behaviour. He asked a large cohort of upper division undergraduate management information systems students to perform a small number of spreadsheet tasks such as building a simple pro forma income statement from a textual description. Students were either assigned to a group of three, or were asked to work alone. Some of the students working alone and in groups were given a warning and told about the percentage of students that had made errors previously while attempting the various tasks. The rest were not told.

After students had built their first spreadsheets, they were asked to estimate the probability that they or their group had made an error when building the spreadsheet. Students working alone and without the warning estimated their error rate to be 18%. Their actual error rate was 86%. The groups estimated that 13% of them had made errors, where actual errors were 27%. Note that working in groups reduces the error rate.

In a second experiment which involved two tasks, the students not given the warning estimated their error rate to be 40%, whereas the students who were given the warning about previous errors estimated their error rate to be 60%, indicating reduced confidence. About 25% of the spreadsheets belonging to students given the warning were correct. Only 7% of the spreadsheets in the groups not given the warning were correct.

People are overconfident about the errors they make when developing spreadsheets. Warning them about the error rate reduces their overconfidence and also reduces their spreadsheet error rate.
Thirty years of scientific research on spreadsheet error has shown that when people develop spreadsheets they make mistakes at about the same rate. The reason they make mistakes is because they are human, and the making of mistakes by humans has been scientifically studied for over a 100 years.

Longstanding research shows that even very simple tasks, such as switching a light bulb on or off according to given instructions, has an error rate of about 0.5%. For more complex cognitive tasks, such as creating spreadsheets, the error rate is around 10 times that rate – around 5%.

The way we measure error rates in spreadsheets varies slightly according to who is doing the research. The central concept however is the Cell Error Rate (CER). Depending upon definitions, this is usually the rate at which unique formulae are found to be incorrect following detailed review (in other research it is based on all formulae). For most research and commercial spreadsheets independently selected for inspection or review, the CER is found to be in the range 1%–5%.

The CER has not varied materially during the 30 years that spreadsheet error research has been going on and includes relatively recent research.

The consequences of the CER are quite profound. If we assume that the CER is 2% for the commercial spreadsheets that we are presently dealing with, then the probability that a spreadsheet cell containing a formulae is correct is 100% - 2% = 98%.

So, for a simple spreadsheet containing 10 formulae, where the formulae are all independent of each other (ie, different) then the probability that all of them are correct is 0.98 to the power of 10, which is around 81%. For a spreadsheet containing 100 formulae, the probability that all of them are correct is 0.98 to the power of 100 or 13% and for 1000 formulae, the probability that all of them are correct is effectively nil.

The research shows that the only way to get around the problem of spreadsheet error (or more rather, human error) is to independently inspect and correct each formula cell. By independent, we mean using people who have not been engaged in the original development of the spreadsheet. The research consistently shows that on each test pass around half of the errors are found.

So, for a 1000 cell spreadsheet with an assumed CER of 2%, we expect that 20 cells will be incorrect. The first formal test of this spreadsheet, involving the inspection of all 1000 spreadsheet cells will result in the correction of about 10 spreadsheet cells. The second independent check of the spreadsheet will eliminate another 5 errors, leaving 5. A third pass through the spreadsheet will leave about 2 errors. Two further comprehensive testing exercises will reduce the likely error rate to less than one expected error. These results are consistent with commercial practice.

The reality is that most organisations will not subject their key and critical spreadsheets to the level of critical review demanded by the rather bald facts as outlined above. This is why the research consistently finds that the majority of spreadsheets used in business are materially incorrect.

And so, there you have it, right at the end of a two year series of blogs, a concise explanation of what the problem with spreadsheets is really about. We all use spreadsheets, we are all human and we make mistakes at a predictable rate. And we don’t (on average) spend much effort managing our natural propensity to make mistakes.

Hence the number one spreadsheet risk is ‘Failure to Review’.
EXCEL ONLINE TRAINING 2014
STRENGTH IN NUMBERS

Excel Community online training is a suite of 38 training modules that consists of a mixture of learning methods, model spreadsheets and video tutorials.

Powered by FILTERED, the modules aim to help users understand Excel better and use it more efficiently, saving them time and money.

From January 2014 the ICAEW Excel Community and IT Faculty will be offering its members new and improved online training as part of their membership.

This is in addition to other benefits provided as part of the Excel Community membership.