Misleading Labels

ABSTRACT

Simplifying complex data into single-value indices can often do more harm than good. This article examines three case studies—a bridge deemed "structurally deficient," a school rated "inadequate," and transformers misclassified by health indices—showing how these reductions can lead to flawed conclusions and unintended outcomes. The discussion emphasizes the need for more comprehensive assessments to avoid critical oversights.

KEYWORDS:

data simplification, evaluation, misinterpretation, complex systems, asset health, risk management

We have three examples of evaluating available data to assess "fitness for purpose" with evaluations which had unintended consequences

Asset health indices have similar issues, whether a grade or a score or whatever: reducing large amounts of data to a single value removes a huge amount of information

As humans, we like to categorize things, put them in "virtual" boxes to help us understand the world around us [1]. It could be as simple as "cars from manufacturer X are unreliable" or "people from Liverpool have a strong sense of humour", and so on. It can be a useful trait to do this, as it helps us make sense of a complex world by applying labels. But it can also be misleading as we tend to assume that everything in a box is the same as everything else in the same box, and everything in the box is quite different to anything outside the box, and this is not always the case [2].

Here, we have three examples of evaluating available data to assess "fitness for purpose" with evaluations which had unintended consequences.

A highway bridge on Interstate 35 in Minneapolis, MN, built in 1967, crossed the Mississippi River, carrying upwards of 140,000 vehicles daily [3]. In 1990, the bridge was classified as "structurally deficient" by the Federal Government due to significant corrosion in bearings. Since 1993, the bridge has been inspected annually by the Minnesota Department of Transport, and several problems have been identified. In 2007, this bridge was one of 75,000 in the USA with a "structurally deficient" classification and was scheduled to be replaced in 2020. On August 1st, 2007, the bridge collapsed, killing 13 people and injuring scores more. A large investigation followed, with many questioning how a bridge rated as "structurally deficient" could be left in service with such a high volume of traffic on it: how bad were the deficiencies and what failure modes were involved over what time scales. There is a lot to "unpack" from the phrase "structurally deficient".
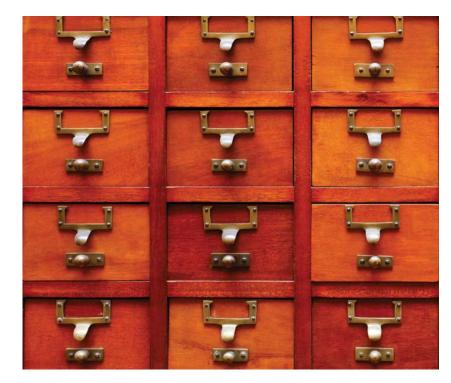
In the UK, the Government body for standards in education and schools is called Ofsted and evaluates schools on a regular basis through on-site inspections [4]. The outcome reduces the school to a one- or two-word grade: "outstanding", "good", "requires improvement" and "inadequate". As one teacher noted: "If a teacher summed up your child in one word, how would you feel?" After one inspection, one school was downgraded from its regular grade of "outstanding" to "inadequate". The Headteacher, who had herself been a pupil at the school, took this very personally and very poorly: she committed suicide, with a subsequent inquest stating her action was a direct result of the evaluation process [5]. The school review system is now being reviewed itself – an aim is to remove the simple categorizations (the one/two-word assessments) as they do not even nearly describe the hugely complex and intricate systems which comprise a school and its operation: a large amount of data boiled down to a one- or two-word summary.

Asset health indices have similar issues, whether a grade or a score or whatever: reducing large amounts of data to a single value removes a huge amount of information. OK, there isn't the same likelihood of someone committing suicide as a result of the evaluation, but there is a similarity in the data "compression". To get value from an index, we need to ensure that the index is meaningful: avoiding what Dr. Harriet Hall of the US Air Force called "Tooth Fairy Science" where numbers are added, subtracted, multiplied and so on, and a number drops out at the end which is supposed to relate to an asset condition [6]. An index is an estimate and thus has a range or precision, and it should also have some basic qualities:

- Monotonic in urgency – a worse index always requires more urgent intervention, and we have a timescale for action
- Calibrated – the same index value for different assets has the same urgency
- Precision – we need to know what the range is on our estimate of condition, and what that means for the urgency
- Auditable – when we plan an intervention, we need to know which data is driving that intervention and why
- Justifiable – we don't do things because the index has a particular value, the index has a particular value because, based on our data and analyses, it reflects our need to do something

One way to think about the worth of an index is to consider whether it would be useful if you only have one asset? What does it do for you? We need to know

The index is a model, and all models are wrong, but how wrong does it have to be before it becomes unacceptably wrong?”

what data goes in, what analyses are performed, and how the results are then used to drive actions: both short-term tactical and long-term strategic. Compressing a lot of data down into a single value and then putting it in a box marked "adequate" means we have removed a lot of information in the data: we have to decide if that is justifiable and acceptable.

In an analysis of >900 transmission transformers, each unit was assessed based on available data, design/manufacturer information, and operational history and placed in one of 4 health index categories: from code 1 being expected to fail within 5 years, to code 4 being "No known fault or design defect". 30 units which were replaced before failure or as a consequence of failure were evaluated to assess the accuracy of the health index process; of the 30 units, 20 were in the category assessed before failure, 5 differed by one category, and 5 differed by more than one category. The actual index was better than expected in 7 cases and worse than expected in 3 cases. This demonstrates the caution we need to use when looking at transformer health indices as they may seem to be extremely precise, but even for world-renowned specialists such as those at National Grid UK, there is always something new to learn, and processes can be improved [7].

Whether we are talking about bridges, schools or transformers, an index is a convenient shorthand for communicating urgency, but it is not the whole picture of asset health. If you think about it, an asset which looks perfectly healthy based on an index is really one where we have not categorically identified any significant deterioration, but it may be that we just haven't looked hard enough! The index is a model, and all models are wrong [8], but how wrong does it have to be before it becomes unacceptably wrong [9]?

References

[1] "Introduction to Human Behavioral Biology", Prof. R. Sapolski, Stanford, https://www.youtube.com/watch?v=NN-nIGh9g6fA

[2] https://recoverynet.ca/2013/11/18/the-perils-of-categorical-thinking/

[3] https://en.wikipedia.org/wiki/I-35W_Mississippi_River_bridge

[4] https://www.bbc.co.uk/news/articles/c4gv0ydj62no

[5] https://www.bbc.co.uk/news/education-67639942

[6] https://skepdic.com/toothfairy-science.html

[7] Jarman P. et al, "Transformer Asset Management: how well are we doing and where do we need to do better?", 82nd International Conference of Doble Clients, Boston, USA, 2015

[8] Box, G. E. P. (1976), "Science and Statistics" (PDF), Journal of the American Statistical Association

[9] Cox, D. R. (1995), "Comment on "Model uncertainty, data mining and statistical inference", Journal of the Royal Statistical Society

Author

Dr. Tony McGrail is the Doble Engineering Solutions Director for Asset Management and Monitoring Technology. He has several years of experience as a utility substation technical specialist in the UK, focusing on power transformer tests and assessments, and as a T&D substation asset manager in the USA, focusing on system reliability and performance. Tony is a Fellow of the IET and a Member of IEEE, CIGRE, IAM and ASTM. He has a degree in Physics with a subsequent PhD in Applications of AI to Insulation Assessment.