FAIR INSTITUTE

Understanding Cyber Risk Quantification
A Buyer's Guide

Created for the FAIR Institute

by Jack Jones, Chairman

Sponsored by: RiskLens®

Table of Contents

Overview

The cyber risk landscape is increasingly impactful, complex and dynamic, and organizations have limited resources to apply to the problem. Furthermore, every dollar spent on cyber risk management is a dollar that can't be spent on other business or mission imperatives. This means that cyber risk management programs must be able to effectively prioritize an organization's cyber risk concerns and choose cost-effective solutions. Both of these require the ability to measure cyber risk well.

Unfortunately, there is a lot of confusion about cyber risk measurement methods — their inherent benefits and challenges, and what the qualities are of "good" cyber risk measurement. As a result, misperception is rampant and it's easy for organizations to leverage methods and solutions that may not fit their needs, or that may be severely flawed. This is true as well for cyber risk quantification (CRQ).

The primary goal of this paper is to help organizations make better-informed decisions about whether CRQ is a good fit and, if so, choose a CRQ solution they can rely on. It will also provide a high-level overview of other cyber risk-related measurement approaches to provide some contrast between them and CRQ.

This paper is organized to approximately align with the process an organization might go through when evaluating and selecting CRQ solutions. Specifically, it:

- Discusses where cyber risk measurement fits within the cyber risk management landscape
- Defines CRQ and describes its value proposition
- Reviews common concerns regarding CRQ
- Provides an overview of other risk-related measurement categories that are often confused with CRQ
- Proposes questions that should be asked of any CRQ solution provider
- Describes red flags that should be looked for in CRQ solutions

Others who may benefit from this paper include industry analysts, consultants, regulators, cyber risk solution providers, and cyber risk technology investors.

Risk management program needs

Risk management programs exist to help their organizations cost-effectively achieve and maintain an acceptable level of exposure to loss. Accomplishing this — particularly within a complex and dynamic landscape like cyber risk — requires being able to do a number of things well:

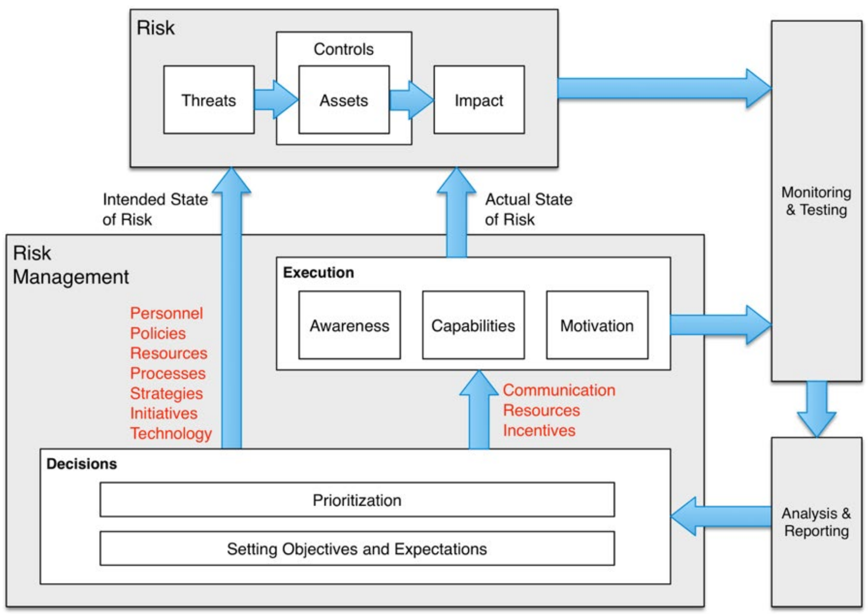
1. Identify/define the potential loss event scenarios that an organization is exposed to
2. Understand the factors (assets, threats, and controls) that affect the probability and impact of those loss event scenarios, and how those factors interact
3. Continually monitor risk factor conditions
4. Given current and projected risk factor conditions, accurately estimate[1] the probability of various loss event scenarios occurring, and their likely impact if they do occur
5. Compare current loss exposure levels against desired states
6. Accurately identify opportunities to reduce risk when or where risk exceeds the desired state, or opportunities to increase risk when or where doing so supports other organization imperatives
7. Accurately and clearly communicate all of the above to appropriate stakeholders so that well-informed decisions can be made
8. Reliably execute the risk management decisions made by executive stakeholders

Although this process may appear to be linear, in reality a risk management program will be doing many of these things constantly and in parallel. Furthermore, because business needs evolve and the risk landscape is always in flux, the monitoring, measurement and reporting aspects of a

1 Because all measurements related to potential future events have some degree of uncertainty they are in fact estimates, regardless of how much or how little empirical data are used in the analysis

program operate as a feedback mechanism. This enables risk management adjustments in response to changes that have occurred to the landscape, or that are expected to occur. Figure 1 below provides a high-level illustration of the risk management landscape.

Figure 1 - Risk management landscape



Given the nature of this landscape, it is clear that the quality of risk management decision-making relies on both the Monitoring and Testing (data) and Analysis & Reporting (analytics) elements of a risk management program.

Although the primary focus of this paper is on CRQ as an analysis and reporting approach, questions related to data also will be discussed.

What is CRQ and how does it help?

CRQ in a nutshell

Well-established risk domains (e.g., credit risk, market risk, operational risk, and various insurance domains such as life, health, and property) have existed for some time to help organizations and individuals manage their exposure to loss — i.e., their risk. In these domains, risk is measured in terms of the probability and magnitude of loss from various scenarios (credit defaults, natural disasters, etc.).

In the cyber risk domain there's rarely any meaningful disagreement regarding risk measurement being a function of loss probability and magnitude. The more common question is what "quantification" looks like and the characteristics of good versus poor CRQ. Later in this paper we'll discuss some of the measurement approaches that are often confused with CRQ. For now, simply recognize that loss event probability[2] is expressed as a percentage (e.g., 10% probability of occurrence in the next 12 months) and magnitude is expressed as a loss of monetary value (e.g., $1.5M). When desired, these values can be combined to express risk as an annualized amount (e.g, $150,000)[3].

It's important to recognize however, that measuring risk quantitatively shouldn't be a goal in itself. What is most important is ensuring well-informed decisions through reliable and meaningful risk measurements (whether qualitative or quantitative).

CRQ value propositions

Prioritization

Being able to recognize which parts of the risk landscape need to be addressed first is fundamental to managing loss exposure well. Historically, cybersecurity has leveraged qualitative measurements to categorize the significance of risk-related concerns. However, even assuming that a qualitative measurement is accurate[4], there remain some fairly obvious questions. For example, imagine that your organization has identified a set of risk-related concerns (Figure 2):

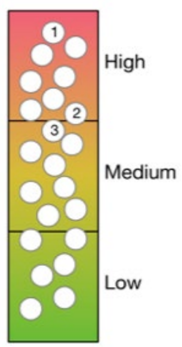


Figure 2

2 Loss event frequency is often a more useful term than probability, particularly with scenarios that can occur multiple times in a year.
3 In reality, to appropriately reflect measurement uncertainty, probability and impact should be expressed as ranges or distributions.
4 Many qualitative cyber risk measurements lack the rigor to be considered reliable.

- How much more risk does the highest "high" (#1) represent than the lowest "high" (#2)?
- In a bucket of "high" risk-related concerns, how do you determine which one is highest and how would you defend that judgment?
- How much more risk does the lowest "high" (#2) represent than the highest "medium" (#3).
- What prevented #3 from landing in the "high" category, and how would you defend that judgment?
- If resources are applied to #1, how much less risk will result and how would you defend those results?
- How much risk is there in total?
- Why were the lines drawn where they were between high, medium, and low risk?

These questions are crucial because organizations need to know where and when to apply their limited resources in order to be most effective. Furthermore, not being able to answer these questions begs the question of whether an organization can effectively defend their risk management choices.

If instead, an organization can rank its risk-related concerns based on quantitative levels of loss exposure, then decision-making and the defense of those decisions becomes easier. That said, there are challenges here as well, which will be discussed in the Common Concerns section of this paper.

Risk remediation cost-benefit analysis

Once an organization has chosen which risk-related concerns to focus on, it now has to figure out what to do about them. With that in mind, if your organization had an additional $1M to spend on cybersecurity, how would it apply those dollars and how much less risk would it have as a result? Conversely, if the organization's CFO knocked on the CISO's door and said that next year the cybersecurity budget would be trimmed by 15%, what would the cybersecurity organization stop doing (or do less of) and how much more risk would the organization have as a result?

The bottom line is that every element within a cybersecurity program — personnel, policies, processes, and technologies — somehow affects the probability or magnitude of the loss event scenarios an organization is exposed to. Understanding which program elements affect which scenarios, and by how much, is foundational to effective risk management and responsible resource stewardship. It is also instrumental in helping executive stakeholders understand the value to be gained (in terms of risk reduction) from cybersecurity investments. This information can only be achieved using CRQ.

Other CRQ value propositions

The ability to prioritize more effectively and choose risk remediation solutions based on cost-benefit analyses is likely to be strong enough justification for many organizations to adopt CRQ. There are, however, other advantages as well, including:

- Senior executives are better able to understand the significance of cyber-related concerns when those concerns are expressed in terms that are similar to other dimensions of their problem space (e.g., economic expressions of revenue, value, operational costs, and other forms of risk expressed in monetary terms)
- The ability to aggregate risk
- Making apples-to-apples comparisons between risks from different domains (e.g., cyber risk vs. market risk, etc.)
- Accounting for cyber risk when considering the value propositions of business initiatives
- Making decisions regarding capital reserves to cover cyber risk exposure
- The ability to select (or price) cybersecurity insurance based on loss exposure
- Meeting evolving regulatory expectations (e.g., SEC guidelines)
- Enabling decision-makers to adjust their decisions when cyber risk measurement uncertainty is faithfully expressed
- Improving the ability to explain or defend risk management decisions.

Despite the apparent value CRQ provides, it has historically faced questions that have limited its adoption. In large part, the hesitance has been due to misperceptions regarding quantitative methods, the sparseness of data, the absence of well-defined models, the lack of CRQ tools, and the inertia that faces any significant change to existing processes. This paper will address the first three of these. Tools are now available, and as adoption continues to grow, inertial resistance will decrease naturally.

Common concerns regarding CRQ

Without question, the two most commonly expressed concerns regarding CRQ are measurement reliability and level of effort. These should, of course, be concerns for any risk-related measurement method, but they come up primarily when discussing CRQ due to various misconceptions.

Let's establish a foundation for discussing these concerns by recognizing several things that happen when an organization chooses to apply its limited resources to prioritize one cyber risk concern over others:

- It reduces by some amount the loss exposure associated with the concern it remediated,
- The resources applied to the concern it remediated are no longer available for any other purpose,
- It accepts (at least temporarily) the loss exposure associated with concerns it didn't remediate, and
- If one or more of the unaddressed concerns contributes to a significant loss event, the organization may have to defend its choices to customers, regulators and other government bodies, investors, employees, or in court.

The purpose of risk measurement is to reduce uncertainty for decision-makers when prioritizing their risk-related concerns and choosing risk mitigation solutions.

Reliability

There are two dimensions to risk measurement reliability — 1) how much faith do we have in the accuracy of a risk measurement method, and 2) even if we trust a measurement method's accuracy, how much does it actually reduce our uncertainty? Looking back at the CRQ Value Proposition section of this paper, the poor performance of qualitative risk measurement in terms of reducing uncertainty is pretty clear. So if qualitative risk measurement is so inherently limited in reducing uncertainty, what are the factors that affect CRQ's reliability?

Accuracy

A common belief in the cybersecurity profession has long been that "You can't quantify cyber risk." The reasons have typically boiled down to some version of:

"There's not enough data.", and

"We face intelligent threat agents that can change their targets and techniques at will, so a risk measurement could be invalidated at any time."

Fortunately, these concerns don't in any way prohibit cyber risk quantification. They do, however, raise the question of whether cyber risk quantification can be relied upon to be (or remain) accurate[5]. It doesn't take a lot of thought however, to recognize that these concerns would apply to any form of risk measurement — quantitative, qualitative, credit-like scoring, etc. Does this mean that organizations simply throw in the towel and apply their resources randomly? Of course not. Despite these challenges, organizations have no viable option but to prioritize the concerns they face as best they can.

With that in mind, the accuracy of any complex measurement[6] hinges on three factors:

- Clarity regarding the scope of what's being measured
- The quality of the analytic model
- The quality of data and how they're applied to the model

The better each of these is within a risk measurement approach (whether quantitative or qualitative), the more accurate the results are likely to be.

In many domains (e.g., life insurance, physics, etc.) measurement approaches can be validated for their accuracy through empirical data or experimentation. That's a much tougher nut to crack in the cyber landscape today, which is discussed elsewhere in this paper. In the absence of those empirical validation options, we can still evaluate the likelihood of risk measurement accuracy using the three factors above.

5 https://www.fairinstitute.org/blog/cure-your-risk-analysis-paralysis-balance-accuracy-and-precision
6 Complex measurement in this context refers to any value (e.g., speed, risk, revenue, etc.) that can't be measured directly and must be derived from two or more variables.

Scope

You can't reliably measure what you haven't clearly defined

One of the most common mistakes in cyber risk measurement is poor scope definition. For example, when an audit finding or vulnerability scan identifies an improperly configured server and the deficiency is rated as "high risk," what exactly has just been measured? Rarely does anyone take the time to identify the specific loss event scenarios a control deficiency or other risk-related concern is relevant to. Poor scoping can take several forms, which feature prominently in the Red Flags section later in this document.

If however, we take the time to identify the specific loss event scenario(s) we're trying to measure the probability and magnitude of, then we can apply an analytic model (e.g., FAIR being one example) and data to measure how much risk a concern like an audit finding represents. The good news is that scoping isn't difficult once you understand the required elements of a well-defined loss event scenario:

- The asset(s) at risk
- The threat(s) that might affect the asset(s) in a manner that results in loss
- The type of outcome from that event (e.g., outage, confidentiality breach, loss of integrity, fraud, etc.)

The more detail you add to a loss event scenario definition (e.g., specific vectors, etc.), the easier it is to apply data effectively, and the higher precision you can have in your results. However, as will be discussed in the Level of Effort section, higher precision comes at a cost.

Analytic models

"All models are wrong, but some are useful.[7]"

This famous quote refers to the fact that all models are simplifications of reality. It also reflects the fact that even though models will always be imperfect, they can be incredibly useful in helping us understand and effectively deal with an infinitely complex world.

There is, however, a continuum of "wrongness" when it comes to models — from relatively minor imperfections to profoundly flawed. In the first case, models will be directionally correct and improve (often dramatically) our understanding of a complex subject and our ability to make well-informed decisions. In the second case, models can be broken in ways that result in poorly-informed decisions. Some of the ways in which models can be flawed are:

- Missing key variables
- Applying variables incorrectly
- Failing to capture dependencies between variables
- Mathematical errors
- Logical errors

These flaws are far more likely to occur in models of highly complex problems (e.g., cyber risk), which is why it's so important for these models to be open for inspection.

Historically, the models commonly used in cyber risk measurement have tended to be extremely unreliable. This is especially true for most qualitative risk measurement, where the "analytic model" is the informal mental model of whoever is proclaiming risk to be high/medium/low. Unfortunately, even some of the formal models currently used in the profession have significant flaws. For example, table G5 in NIST 800-30 has a significant logical flaw that allows the overall likelihood of loss to be higher than the likelihood of events that would create the potential for loss[8].

Formal analytic models in any measurement discipline can be developed either inductively using empirical data that "informs" us about how the world seems to work, or deductively by carefully decomposing a measurement objective (e.g., risk) to identify key factors and their relationships. This is how Factor Analysis of Information Risk — FAIR — was developed. For those readers with deeper statistical backgrounds, this deductive approach is essentially Bayesian in nature. Both approaches are valid, and both have advantages and disadvantages that go well beyond what can be discussed in this paper.

What's important to recognize is that to-date there have not been sufficient empirical data to create reliable CRQ models inductively for the majority of the cyber risk landscape. This situation will improve as data quality and quantity improves, but for now it's relatively safe to assume that more robust

7 https://jamesclear.com/all-models-are-wrong
8 https://www.fairinstitute.org/blog/fixing-nist-800-30

and broadly applicable cyber risk measurement models have been developed deductively. The question then becomes how solid the deductive effort has been, and how thoroughly a model has been exposed to examination for logical errors[9].

Data

> *"You have more data than you think you do, and you need less data than you think you do.[10]"*

Even though scoping and analytic models affect measurement accuracy at least as much as data does, the concern most commonly voiced regarding cyber risk measurement is about data quantity and quality.

A useful question to ask when people express concerns about data is, "How much data would be enough?" The most common responses are either a blank stare or something along the lines of, "Something statistically significant, similar to what's used in insurance actuarial tables." A blank stare suggests they're merely regurgitating something they've heard but not thought much about. An insistence on statistical significance suggests they don't recognize at least three facts:

- Significant volumes of data don't exist for much of the cyber risk landscape, due to a combination of relatively low loss event frequency for many events, the lack of data sharing, and highly variable conditions from organization to organization.
- Qualitative risk measurements also are based on data; how else would someone justify "high" versus "medium"?
- There are well-established methods for reducing and accounting for data-related uncertainty.

Every complex discipline — even a "hard science" like physics — has to deal with measurement uncertainty. The key is to effectively account for the uncertainty in your inputs using ranges or distributions, and apply methods like Monte Carlo[11] functions to account for uncertainty when applying math, and to reflect uncertainty in results using distributions, error bars, etc. With risk measurements that faithfully and realistically reflect uncertainty, decision-makers are in a much better position to do at least two essential things:

- When measurement uncertainty is high, they can be more conservative in their decisions, and
- They can allocate additional resources to reduce uncertainty if and when desired.

One of the significant advantages of CRQ over qualitative or other forms of risk-related measurements is that (when well designed) it enables faithful representation of measurement uncertainty, which is fundamental to making a well-informed decision.

There is another important point to make regarding data. Even though the lack of data doesn't prohibit reliable measurements (as long as uncertainty is accounted for), how data are used in an analysis matters a lot. There are several concerns regarding this in the Red Flags section of this paper, that highlight common data-related mistakes. Another white paper[12] discusses in more detail the challenges associated with appropriately using cybersecurity data.

Given the above discussion, persons claiming that "CRQ is just guessing" are either referring to poorly designed and executed risk measurement, or aren't well-informed on robust quantitative methods. Furthermore, because concerns regarding accuracy apply to any risk-related measurement, the question of CRQ accuracy has to be cast within the context of, *"Compared to what?"*

Uncertainty reduction

Assuming that a measurement's accuracy isn't in question, there remains the question of how much it reduces a decision-maker's uncertainty about risk. This might seem to be an odd statement, so perhaps a couple of examples will help get the point across.

- How much does a NIST CSF score of "3.2" reduce uncertainty regarding how much risk an organization has, or how much less risk it will have if one or more CSF elements is improved?
- How much does a cybersecurity credit-like score of "682" reduce uncertainty regarding how much risk an organization has, or how much less risk it will have if improvements are made?

Even if both measurements were arrived at logically and without significant errors (i.e., are "accurate"), neither one measures risk in terms of the probability or magnitude of loss. Consequently, both of them are inherently limited in

9 This is the primary reason why the FAIR model was made an open international standard.
10 From "How to Measure Anything" (by Douglas Hubbard)
11 Monte Carlo and other stochastic methods were developed specifically to deal with uncertain data.
12 https://www.fairinstitute.org/blog/white-paper-effectively-leveraging-data-in-fair-analyses

their ability to reduce decision-maker uncertainty regarding risk levels. We may be willing to assume that higher scores imply less risk, but we don't know how much less risk. Nonetheless, these types of relative scores can still be very useful in reducing uncertainty about whether a cybersecurity program is improving or degrading, and where it stands in relation to other programs (i.e., benchmarking).

Another consideration regarding uncertainty reduction takes us back to something that was touched on in the Data section above — the question of precision. For example, we can guarantee accuracy in a risk measurement by stating that the probability of loss is somewhere between 0% and 100%, and the impact will be somewhere between $0 and infinity. That "measurement" required no effort at all and there's very little chance of being wrong, but it hasn't reduced uncertainty in any way because it doesn't offer a useful degree of precision. We can work harder to provide narrower ranges (i.e., better precision) but that takes time. The point is, measurement precision affects how much uncertainty is reduced, and it also plays a significant role in the cost of risk measurement.

Level of effort

How much effort an organization puts into CRQ is highly dependent on two factors that are largely within the organization's control — measurement comprehensiveness and precision.

Within a cyber risk context, comprehensiveness is a question of how much of an organization's cyber risk landscape it wants to measure. Although the obvious answer might seem to be "All of it, please.", it's fairly obvious that not all parts of the cyber landscape are equally important. For example, by definition an organization's crown jewels have higher value/liability characteristics than other assets and therefore probably deserve greater attention.

But comprehensiveness involves more than just the assets at risk. It also includes various the types of loss events (e.g., outages, information compromises, data integrity loss, financial fraud, etc.), different threats (e.g., trusted insiders, third parties, cybercriminals, nation-state actors, technology failures, acts of natures, malicious, human error, etc.), and different kinds of controls.

In a perfect world, we would have copious cybersecurity telemetry for all of the many loss event scenario permutations these variables represent. If this were the case we would have a source of truth that solves both the comprehensiveness and precision concerns. However, that isn't our reality today. This means that for significant parts of the cyber risk landscape, organizations have to work with sparse data. Fortunately, other disciplines (e.g., various sciences) have developed very effective methods for dealing with sparse data and measurement uncertainty.

Consequently, the keys to managing an organization's CRQ level of effort boil down to:

- Knowing which parts of the risk landscape it wants to focus on, and
- Knowing when and how to leverage more precise data (e.g., telemetry) versus less precise data (e.g., SME estimates)

Another potential point of leverage is to leverage CRQ technologies to scale an organization's risk measurement efforts.

A number of vendors have begun producing CRQ solutions that make various claims regarding ease of use and simplicity. Some of these claims are justified, and the methods for achieving them maintain analysis reliability. Other claims, however, are based on methods that significantly compromise the quality of analysis results. The Red Flags section of this paper provides guidance about the most common problematic shortcuts.

At the end of the day there is no single correct answer to the question of how much effort is required to incorporate CRQ into an organization's program, as every organization will have different needs and resources. What's important is for organizations to recognize that there is flexibility in this regard, and that they have have significant control over this concern.

Other risk-related measurement approaches

As mentioned earlier, other approaches to risk-related measurement exist, and it's useful to recognize the value they bring to the table, as well as their limitations. For the record, even though these approaches provide numeric results, none of them quantify and express the probability and magnitude of cyber-related loss in financial terms. Consequently, these methods do not support the primary CRQ value propositions, nor do they qualify as CRQ.

That said, individual vendor solutions within some of these measurement approaches may take steps to bridge the gap between their primary capabilities and CRQ. This can be a good thing or a bad thing, depending upon how carefully they've done their homework. Just like any other solution that claims to do CRQ, such hybrid solutions should be evaluated using the Important Questions and Red Flags sections of this document.

Numerically expressed ordinal risk measurement

One widespread approach to risk measurement has been to use simple 5x5 (or similar) scales for probability and impact. Typically, these scores are then multiplied to arrive at a "risk score" (e.g., 2x5=10). GRC tools, spreadsheets, and even some risk measurement software solutions often use this approach.

Although this approach uses numeric values, we have to recognize that 1-thru-5 (and similar) scales are ordinal values that can be replaced with words (e.g., High, Medium, Low) or colors (e.g., red, yellow, green). Therefore, numeric scales like these are not quantitative values but instead represent labels for "buckets" that permit high-level grouping and ordering (thus the term "ordinal").

Capabilities

If the goal is to categorize or group one or more potential loss event scenarios (e.g., an outage due to accidental, malicious, or natural act, etc.) based on ordinal likelihood and magnitude scales, then this can be a quick and relatively low-cost approach. However, even when operating at this simplistic level, it is still crucial to properly scope the loss event scenario(s) that you're estimating probability and impact for.

Limitations

These measurements are ordinal values (i.e., labels), which means that although we might believe (for example) that a loss event scenario rated as a "1" for likelihood is less likely to occur than an event rated as a "2" for likelihood, we have no way of knowing how much less likely. We also don't know whether the difference between a 1 and a 2 is the same as the difference between a 2 and a 3, etc. What this means is that performing math on these values is unreliable (at best). A useful article on measurement scales and their uses/limitations can be found here[13] and a more detailed analysis of the limitations of ordinal methods can be found here[14].

A related problem with ordinal scales is that definitions for each level in the scales are often just descriptive verbiage, like: "Low Probability = Unlikely to occur." One challenge with this is that verbiage such as "unlikely to occur" or "significant impact" are open to subjective interpretation[15], which means two people rating the same thing can easily end up with two different answers. Another common deficiency is that the probability scale descriptions often don't time-bound the description. For example, "Unlikely to occur" in what time frame? This year? This month? In our lifetime? Without being time-bound, these descriptions are even less reliable as measurements.

Because ordinal scales don't provide clarity regarding whether, or to what degree, differences from one level to another are similar, it is highly unlikely that applying math to them will provide reliable results that stand up to scrutiny. There is one possible exception, and that is if the points in the scale (e.g., 1, 2, etc.) represent non-ordinal ranges rather than subjective terms like "significant loss", "almost certain", "unlikely to occur", etc. For example:

Value	Probability (next 12 months)
5	90% - 100%
4	70% - 89.9%
3	30% - 69.9%
2	10% - 29.9%
1	0% - 9.9%

Value	Impact
5	> $10M
4	$1M - $9.999M
3	$100k - $999k
2	$10k - 99k
1	< $10k

With scale definitions like these you could, arguably, apply math on the ranges within the scales. For example, if you said the probability of a particular loss event was a "2" in your scale, and the impact of the event was expected to be a "5", you could multiply the ranges represented by those values (e.g., 10% x $10M and 29.9% x $10M) to arrive at an annualized loss exposure level of somewhere between $1M and $2.9M. It works, but there are still several obvious and significant problems, including:

- How do you choose an upper-bound for the highest range in the level of impact? If you don't provide a range for the highest level, what monetary value do you use in your calculation? If you simply used $10M, you would not effectively capture or reflect losses that may be much larger.
- What if the loss event scenario being analyzed could occur multiple times in a year? The highest probability range won't effectively represent that condition. In this case, you would be better off using a frequency scale.
- Do the selected ratings (e.g., 1, 2, etc.) represent best-case, worst-case, or most likely case? This has to be established in order for the results to be clearly understood, defensible, and appropriately applied when making decisions.

13 https://www.mymarketresearchmethods.com/types-of-data-nominal-ordinal-interval-ratio/
14 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.4544&rep=rep1&type=pdf
15 https://hbr.org/2018/07/if-you-say-something-is-likely-how-likely-do-people-think-it-is

- What do you do if you're unsure whether the probability or impact is in one range versus another, or spans more than one range — e.g., the probability of an event is believed to be between 15% and 50%. Do you combine both ranges, making it 10% to 69.9%?

- How do you decide on the range boundaries? This is particularly challenging for the Impact scale if your organization is made up of divisions or subsidiaries of different sizes that have different loss magnitude tolerances.

An approach like this may get past the ordinal math problem, but it introduces as many challenges as it solves. Furthermore, if you know enough about the loss event scenario you're measuring to select from a set of predefined ranges, there is no reason why you can't avoid these pitfalls altogether and take the next logical step, which is to eliminate ordinal scales and use data and calibrated estimates for your measurements.

Controls-focused assessments

These are sometimes also referred to as risk management maturity assessments. Examples include: NIST CSF[16], NIST 800-53, PCI DSS, ISO2700x, COBIT, etc. Note that the line between maturity model assessments and controls-focused assessments can be thin and blurry, and it isn't uncommon to see one referred to as the other or contain elements of the other.

These frameworks provide a list of controls or control outcomes that are considered important (or common) enough to test when evaluating a cybersecurity program. They can be made up of elements that are relatively high-level in nature (e.g., NIST CSF) or much more granular (e.g., NIST 800-53). Scoring can be binary (the control exists or not) or more nuanced using a scale to reflect a degree of coverage, efficacy or maturity. These scales are almost always ordinal (e.g., red/yellow/green, 1 thru 5, etc.).

Capabilities

Controls-focused assessments are essential tools for evaluating whether specific controls are in place or working as intended. This provides a crucial piece of information for effectively managing a cybersecurity program. This information also can be useful within CRQ analyses when combined with other data points, such as threat event frequency and loss event impact.

Limitations

Despite their value in identifying control deficiencies, control assessments don't measure risk and therefore, by themselves don't provide the "So what?" behind a deficient finding. For example, we can't know how to prioritize one control deficiency versus another without determining the risk implications through reliable risk measurement. Unfortunately, very often these prioritization decisions are made without careful consideration of some crucial facts:

- Control assessments are measuring control conditions — not risk. The relevance of a deficient control depends on the value/liability characteristics of the asset(s) it's protecting, as well as the threats it's protecting against, and (as described in more detail below) the condition of other controls. Consequently, measuring a control's value proposition can only be determined through quantitative risk analysis.

- Not all controls are of equal importance, either overall or within the context of specific loss event scenarios. To drive this point home in a conversation on this topic you can ask the question, "Which is more important, authentication or logging?" Most of the time, the answer will be "authentication" under the premise that an ounce of prevention is worth a pound of cure. The problem is that if the loss event scenario we're trying to measure and manage involves a malicious privileged insider (e.g., a rogue network administrator), then authentication is a moot point because we've intentionally given the person access. In that case, logging is the more critical control.

- Another challenge is that none of the control frameworks today take into account the relationships and dependencies between controls. For example, logging and monitoring have a Boolean[17] "AND" relationship with one another because both have to be working in order for the organization to fully realize their intended value. In fact, every control has a relationship with at least one other control, and sometimes there can be many relationships. None of the standard control frameworks take this into account, which means that scoring control efficacy using these frameworks is incomplete at best and can often be grossly inaccurate.

- When control conditions are measured using ordinal scales, the use of math on those values has to be considered unreliable. You cannot, for example, average your control condition scores and expect the results to stand up to scrutiny, especially given the challenges described in the previous two bullets.

16 NIST CSF describes itself as a controls outcome framework, which in practice isn't different enough to create a separate category in this document.
17 https://hsl.lib.umn.edu/biomed/help/boolean-operators

Vulnerability assessments

It is common practice in today's cyber risk management landscape for organizations to use scanning technologies to identify technology-related weaknesses in their defenses (missing patches, improper configurations, poor software design, etc.). Most often, these technologies leverage the Common Vulnerability Scoring System (CVSS) or a derivative of that model to rate the significance of any findings.

Capabilities

These technologies are excellent at identifying technology-related weaknesses, which is crucial information for managing cyber-related risk.

Limitations

Although CVSS-based tools are quite capable of identifying weaknesses, the CVSS scoring model does not measure risk, let alone measure risk quantitatively. In fact, CVSS scores only capture a fraction of the information necessary for accurately measuring risk. It does not include crucial parameters that help organizations understand the frequency of attacks or the impact of loss events. In addition, the CVSS scoring model suffers from several of the problems highlighted in the Red Flags section of this document, including incomplete scoping, math on ordinal scales, and the use of weighted values.

Credit-like scoring

In this approach, an index value (the credit-like score) is created from various data points, which may include some of the organization's control conditions, data traffic patterns, characteristics of the organization's industry (value/liability considerations), industry threat-related data, and perhaps even some industry-related loss event factors. These data points are fed into an algorithm that generates the score.

Capabilities

Because these solutions tend to be consistent in terms of how they use data and the algorithms being applied, this is currently the most reliable means of benchmarking one or more organizations against others. This can make it a very useful tool for evaluating the security condition of third-parties an organization does business with, as it can be excellent at identifying specific types of control deficiencies (e.g., missing website encryption, software patching, etc.).

In addition, an index score can be easy for boards and executives to relate to, and answers two questions they commonly have:

- Where do we stand relative to others in our industry?
- Are we improving (or backsliding)?

Another upside is that these solutions tend to operate relatively automatically, leveraging complex algorithms and working almost purely from data sources like scanning technologies, traffic analysis technologies, etc. When done well, this can be especially valuable for providing a relative ranking of third parties, and to recognize profoundly deficient third-party control conditions.

Limitations

Although credit-like scoring approaches can be beneficial for the purposes mentioned above, it's also important to recognize what they can't provide. Specifically:

- They are not CRQ solutions because they do not measure how much risk exists. For example, a score of 742, implies that less risk exists than a score of 581, but we don't know how much less loss exposure that difference represents.
- Historically, these solutions have only had access to internet-facing data (SSL certificates, scan results, traffic analysis, etc.) with which to evaluate an organization, and their algorithms assume that those data points reflect how an organization cares for all of its systems. Consequently, a significant part of an organization's risk landscape may not be included in, and may be misrepresented by, analysis results.
- An index score boils a vast ocean down to a single score. Granted, these solutions often break down their results into various sub-categories, but many of the risk management decisions an organization needs to make (prioritizing audit and other security-related concerns, processing policy exception requests, etc.) cannot be accomplished using these scores because they don't directly measure risk. Similar to averages (which don't disclose key data points like best-case, worst-case, etc.), important information may not be revealed through index scores.
- Measurement uncertainty is not expressed in these scores even though the data feeding these analyses is invariably imperfect, and significant modeling assumptions underlie the analyses.

Maturity model assessments

Formal approaches to maturity models (e.g., CMMI, etc.) typically use an ordinal scale of 1 thru 5 or 0 thru 4 to reflect different levels of maturity for specific processes within a risk management program or for the program overall.

Capabilities

Well-designed maturity models can be an important tool in a risk management program, by providing two main benefits:

- They are useful for identifying risk management process deficiencies
- They can be effective for tracking and reporting process improvements over time

Limitations

Although maturity models tend to be numeric, they don't qualify as CRQ because they measure risk management process or program maturity — not risk. Consequently, a maturity assessment does not describe how much risk an organization has and can't be used to prioritize deficiencies or perform cost-benefit analyses of proposed improvements.

Because these values are ordinal, they suffer from the same limitations concerning math — i.e., the results of any math performed on these values are unlikely to be reliable.

Another point to keep in mind regarding maturity models is that they are designed to gauge the maturity of processes. So if you want to use a maturity model that leverages the standard maturity model scale definitions, you need to ensure that what's being measured are processes as opposed to non-process controls such as firewalls, logging technologies, etc.

Threat analysis

There has been a growing recognition of the need to analyze and better understand the cyber threat landscape. As a result, threat analysis models have been developed over recent years to formalize and improve the ability to evaluate an organization's threat landscape. Two of the better-known models are DREAD[18] and STRIDE[19].

Capabilities

Threat-focused models enable a much richer and more reliable understanding of the threat landscape and how various malicious events can transpire. This information can be exceptionally useful for identifying and mitigating vulnerable conditions in software, technology architecture, and processes. It also can provide valuable insights into threat event frequency and threat capability variables when performing CRQ analyses.

Limitations

Threat-focused models tend to focus primarily on the threat and vulnerability components of a risk analysis, which means they often leave out other critical risk factors. They also tend to be exclusively focused on malicious cyber events. As a result, organizations relying solely on these models are only covering a portion of their cyber risk problem space (e.g., ignoring losses due to technology failures, acts of nature, and human errors like mistakenly sending sensitive information to the wrong recipient). Another significant limitation is that these models tend to rely on ordinal measurement approaches — including using math on ordinal values.

Important questions

Although many of the questions in this section can be applied to any risk-related measurement, in this paper they have been specifically written to focus on CRQ.

When evaluating CRQ solutions, organizations should consider including the questions described in this section to their evaluation criteria. This list is not comprehensive, as each organization may have additional areas of interest or concern. Nonetheless, questions from this list should serve as a good starting point.

Questions regarding utility

How does a CRQ solution define the "risks" being measured?

If there is a "most important" question to ask any CRQ provider it is this one, because if risk scoping is done poorly, their measurements will not be reliable regardless of the data they use or the algorithms they apply. The bottom line is that the only things you can assign a probability and impact to are loss event scenarios. Note that there are several common problems related to this in the Red Flags section of this document.

18 DREAD stands for Damage, Reproducibility, Exploitability, Affected Users, and Discoverability
19 STRIDE stands for Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privileges

What part(s) of the risk landscape does their solution analyze?

Cyber risk is a subset of the overall technology risk landscape, which is a subject of the broader operational risk domain. Very often, the lines between these domains are blurry, and CSO's may have responsibilities that cross these boundaries. Consequently, it's necessary to understand what can and cannot be measured with a particular CRQ solution. The following are examples of risk landscape components that may or may not be within the scope of some CRQ solutions:

Assets

- Data?
- Systems?
- Facilities?
- Personnel?

Threats

- Human error?
- Acts of nature?
- Malicious outsiders (cybercriminals, nation-state actors, etc.)?
- Malicious insiders?
- Technology failure?
- Competitors?

Vectors (for malicious actors)

- Attacks thru technology (e.g., network vulnerability exploitation, etc.)?
- Attacks thru personnel (e.g., phishing, etc.)?

Types of outcomes?

- Operational outages?
- Customer information breaches (PII, PHI, etc.)?
- IP disclosure/theft?
- Fraud (direct theft of money)?
- Regulatory compliance failures?

Questions related to data

There are a lot of misconceptions regarding the challenges and opportunities associated with security telemetry and risk analysis. The lack of normalized data, the ambiguous nature of much of the data, and poor scoring methodologies used by many cybersecurity technologies are just some of the challenges. Also, control efficacy data available today are usually seriously flawed, which is described in the Red Flags section of this document.

The questions in this section (and the Analytic section that follows) are intended to help you understand the data a CRQ solution may be using and how those data are applied.

What data do they use for the frequency of attacks/events?

- Data from specific threat-related technologies, such as; IDS systems, firewalls, server and application logs, SIEM solutions, etc.?
- Data from sources such as Verizon DBIR, etc.?
- Information sharing forums, such as the Information Sharing and Analysis Centers (ISAC)?
- Internal subject matter expert estimates?
- Some combination of the above?

One of the fundamental components in any risk measurement is the probability of a loss event. Unless an organization has empirical data based on actual loss experience (which is not typical for many cyber scenarios today), it is necessary to estimate loss event probability directly or derive it from a combination of threat event frequency and control efficacy values.

Where does control-related data come from?

- GRC tools?
- Controls framework assessments (e.g., NIST CSF, FFIEC CAT, etc.)?
- Vulnerability/configuration scanning technologies?
- Anti-malware technologies?
- Internal audit and other self-assessments?
- Industry data such as Verizon's DBIR?
- Internal subject matter expert estimates?
- Some combination of the above?

Control-related data comes in many forms and from various sources. This variability is both a blessing and a curse. On the plus side, variety can add robustness to data. On the negative side, it adds significantly to the difficulty in correctly using data. Note that there are several red flags regarding control-related data later in this document.

Where does impact data come from?

- Verizon DBIR?
- Commercial firms that do impact research?
- Cyber insurance providers?
- Ponemon studies?
- The CRQ provider's own research?
- Internal subject matter expert estimates?
- Some combination of the above?

Historically, reliable impact data have been sparse, and much of the data that has been published hasn't been well-researched. Making matters even more challenging, some forms of loss have evolved rapidly, which means the useful lifetime of historical data can be shorter. Fortunately, more impact-related data are becoming available, although at the time this paper is being written we're still a long way from having high volumes of reliable loss data.

Is impact data cleanly broken out into categories?

- Lost revenue?
- Response costs?
- Replacement costs?
- Monetary loss thru fraud/theft?
- Reputation damage?
- Fines/judgments?
- Others?

Another challenge associated with impact-related data is the absence of a universally agreed-upon taxonomy for gathering or categorizing it. The closest such standard today is the one defined by FAIR. Regardless of what taxonomy is used, decomposing the different ways in which loss materializes will reduce the odds of missing data or double-counting data.

What assumptions do they make when applying historical data?

This question is vital for all CRQ solution providers but is particularly crucial for solutions that rely heavily on security telemetry or other sources of risk-related data in an attempt to reduce the need for subject matter expert judgment. Specific concerns in this regard include:

- Given the highly dynamic nature of the cyber risk landscape, how do they reflect the degree to which historical data may not reflect the future?
- How do they account for ambiguity in threat data? Many threat-related data points (e.g., authentication and access logs, etc.) do not indicate whether an event was malicious or accidental, or if malicious in nature, whether the intent was to steal data (of what kind), bring down a system, etc. These considerations can make a significant difference in the accuracy of risk analysis, and are an example of why attempts to "eliminate expert judgment" often generate unreliable results.

- How do they account for differences in impact data based on the specific industry and business model of your organization? For example, do the impacts primarily reflect retail-based organizations, banks, healthcare, manufacturing, non-profit, or does it allow you to customize for your specific organization?

How is measurement uncertainty accounted for?

- Predefined ranges for inputs?
- Ability to reflect best-case, worst-case, most likely case for inputs and outputs?
- Ability to shape input distributions (e.g., skew)

Faithfully accounting for measurement uncertainty in the inputs (and outputs) of a CRQ solution is crucial in order for the results to stand up under scrutiny and provide reliable information to decision-makers.

How is fallout from an event (e.g., fines & judgments) analyzed?

Many loss event scenarios have the potential for additional (often severe) fallout forms of loss, such as fines, customer churn, etc. How a CRQ solution deals with fallout is a crucial factor in generating accurate results.

How are data and/or analyses updated as the risk landscape evolves?

As the risk landscape evolves, old risk analysis results may no longer accurately represent an organization's loss exposure. With this in mind, it is important to understand how a CRQ solution enables updates to previous analyses.

What distribution types are they using?

Those readers with deeper backgrounds in statistics will recognize the difference between things like Normal distributions, Uniform distributions, Poisson distributions, Power Law distributions, Beta-PERT distributions, etc. What is important to recognize is that some distributions are better suited than others for certain risk-related data points. When data are sparse and calibrated subject matter estimates are more heavily leveraged, Beta-PERT distributions are commonly used. When data are plentiful (which is still relatively unusual for much of the cyber risk landscape) then the most appropriate distribution type for a given situation will be easier to discern. Regardless, you will want to understand the choices a solution provider has made for distributions, and the reasoning behind those choices.

Questions related to analytics

Does their algorithm perform math on ordinal values?

Ordinal values (e.g., 1 thru 5 scales, high/medium/low, etc.) are commonly used within overly simplistic risk measurement solutions. However, because of the fundamental nature of ordinal values, any math performed on them should be considered unreliable.

How is control efficacy measured and represented?

- Ordinal values?
- Maturity ratings?
- As a percentage?
- Something else?

This is one of the most challenging and commonly flawed capabilities in CRQ solutions, as discussed within the Red Flags section of this document.

Does it support what-if analyses?

One of the most powerful capabilities of a good CRQ solution is being able to do what-if analyses to reflect the potential cost-benefit of proposed risk management improvements, or to reflect the potential downsides of control degradation, changes in the threat landscape, etc.

How has their model been validated?

If the underlying analytic model is fundamentally flawed, then it doesn't matter how good your data are — the results will not be reliable. There are two ways of validating any risk analysis model:

- If sufficient empirical data exists, you can back-check a model by applying historical data to the model to see how closely it generates results that match what has been experienced over time.
- If sufficient empirical data does not exist for back-checking, you can validate a model by ensuring that it doesn't violate known measurement principles (e.g., math on ordinal scales, failing to reflect uncertainty, etc.), that it stands up logically (e.g., that it includes all of the necessary components of a risk measurement), and that it leverages methods that are proven in other measurement domains (e.g., Monte Carlo, etc.).

Back checking is typically preferred when it's feasible, but today there aren't enough empirical data within most of the cyber risk domain to reliably validate CRQ models in that manner. Consequently, the second validation approach is usually a more realistic solution.

Ideally, independent validation has been performed by a standards organization or other reliable and independent source, but in the absence of this you can leverage the Red Flags section of this paper to evaluate CRQ solutions and minimize the odds of adopting a seriously flawed solution. It's also important to point out that model validation is (or should be) a concern for any risk analysis method, whether qualitative, quantitative, credit-like scoring, etc.

Questions related to reporting

How are the results expressed?

- Discrete values?
- Distributions of probabilities?
- Loss exceedance curves?

Recall from earlier in this paper that in order to qualify as a CRQ solution, results have to be expressed in monetary loss exposure terms.

Does it break out results into probability and magnitude components?

Loss exposure is invariably a function of the probability and magnitude of a loss event. Therefore, it is useful to be able to see these values separately so that you can better explain results to stakeholders, and for use as KRI's. This also can be important when considering risk mitigation measures or other changes to the risk landscape (e.g., threat landscape changes, etc.) which might affect probability but not magnitude (or vice versa).

Does it aggregate risk?

Understanding how much risk exists for a specific loss event scenario or control deficiency can be useful for tactical decision-making, but for many strategic risk decisions it's important to know how much aggregate risk exists for the organization overall, or for multiple scenarios.

Is it able to report risk levels by organization unit, line of business, etc.?

Not every part of an organization introduces the same types or levels of risk. Consequently, it can be essential to recognize where higher or lower levels of risk are coming from.

Does it provide cost-benefit results?

Some of the most important risk management decisions an organization can make will be choosing between risk mitigation solutions. In order to do this well, an organization needs to be able to compare the risk reduction value proposition and cost implications of various mitigation options.

Does it report risk level trending?

Because the risk landscape evolves over time, it can be useful to display how an organization's level of risk (or certain other risk-related values) change over time.

Also, because your organization may have its own tracking and reporting requirements, it is useful to know whether a CRQ solution allows you to choose tracking periods (e.g., quarterly, monthly, etc.).

Does it enable reporting against risk appetite?

Achieving and maintaining an acceptable level of risk requires that an organization define and measure itself against a risk appetite. A CRQ solution should enable this capability in some meaningful way.

Can it export to different formats?

Risk analysis results often are used in PowerPoint presentations, Word documents, or may be further analyzed using statistical analysis and reporting tools such as Tableau. Exporting analysis outputs, input values and sometimes even intermediate values (e.g., individual Monte Carlo simulation values) may also be useful for additional analysis.

Common export formats include Excel, CSV, Word, or PDF documents.

How does it reflect uncertainty in results?

Because risk measurement inputs almost always have some degree of uncertainty in them, this should be faithfully reflected in the output so that executives have the opportunity to adjust their decisions accordingly. Typically, the uncertainty in risk inputs and outputs are expressed as ranges, distributions, probabilities, using error bars, etc.

Red flags

When looking for a CRQ solution there are some red flags to watch out for in solution provider claims and approaches. These red flags can indicate something as benign as marketing hyperbole, or as dangerous as profoundly inaccurate analytic results. Either way, you need to know how much you can rely on and defend the results from a CRQ solution before you use it.

With that in mind, here are some of the most common characteristics and claims that deserve significant due diligence.

"Risks" that aren't loss event scenarios

Significance: If risks are defined poorly, no amount of data or math is going to generate reliable results.

Discussion: Risk is almost universally measured in terms of the probability of a loss event scenario occurring, and the magnitude of loss that is expected to result if it occurs[20]. The point to keep in mind is that these two variables can only be applied to loss event scenarios.

Unfortunately, it's common to see people and solution providers confuse and conflate the things that make up our risk landscape (e.g., control deficiencies, threats, assets, etc.) with risks. For example, you'll often see things like these in a list of "risks":

- Weak passwords
- Disgruntled insiders
- Out-dated technologies

It isn't hard to see that although these all contribute to how much risk an organization has, they are very different from one another (respectively — a control deficiency, a threat community, and a category of assets) and none of them are loss event scenarios. For example, we could combine these three elements to define an actual loss event scenario that can be measured using probability and impact:

"A disgruntled insider leverages weak passwords on outdated technology to steal sensitive corporate information."

Weak passwords, disgruntled insiders and outdated technologies could each apply to many other loss event scenarios as well, which means that trying to assign an accurate probability and impact to these concerns individually is not feasible. Nonetheless, you will encounter CRQ solutions that don't recognize this fact and try to apply probability and impact values to things that aren't loss event scenarios.

A variation of the "risks that aren't loss event scenarios" problem are incomplete descriptions of what's being measured. An example taken from one CRQ solution is:

"Attackers exploit weak default configurations of systems that are more geared to ease of use than security."

In this example, we don't know whether the attackers are internal or external (the probabilities and impacts can be very different), nor do we know whether the event resulted in an outage, a confidentiality breach, or some other outcome (the impact can be dramatically different across these). As a result, any measurement of that "risk" should be considered unreliable.

Scope overlap

Significance: Without careful scoping, risk aggregation will not be reliable.

Discussion: Another significant problem in some CRQ solutions is a failure to clearly define the scope of the loss event scenarios being measured. Examples taken from one CRQ solution are:

"Attackers exploit and infiltrate through network devices whose security configuration has been weakened over time by granting, for specific short-term business needs, supposedly temporary exceptions that are never removed."

"Attackers are able to gain unauthorized access to systems due to gaps in security governance and/or enforcement of security policies."

Besides both of these suffering from the "incomplete description" problem discussed in the previous red flag, they also clearly overlap one another. The first scenario is a subset of the second one.

CRQ solutions have to define the risks they measure in a mutually exclusive manner in order to avoid duplication and perform reliable aggregation.

20 There are exceptions, such as when someone tries to measure and articulate "positive risk," which isn't relevant to this paper's focus.

Ordinal measurements

Significance: With few exceptions, any solution that performs math on ordinal values should be considered unreliable.

Discussion: If any of the inputs being used in a CRQ solution are ordinal (e.g., 1, 2, 3, red, yellow, green, high, medium, low, etc.) you need to know what those values represent and how they're being used. Too often, these values and the math performed on them will not stand up to scrutiny and can't be defended. The problem with math on ordinal scales is well-documented[21], but one excellent resource to delve deeper in this topic is Douglas Hubbard's book, The Failure of Risk Management.

One of the most common places to see ordinal values being used in CRQ is for control effectiveness or maturity. If you encounter this, you should ask what a "3" (or whatever) represents, how much different it is from a "2", and whether a "3" for a preventative control is equivalent in its effect to a "3" in a recovery control. It is unlikely that the answers you receive will demonstrate a clear understanding of the challenges associated with ordinal measurements, or stand up under close examination.

A possible exception would be if the ordinal values map to predefined quantitative ranges (e.g., percentages, frequencies, dollar amounts, etc.) or distributions. Even then, you need to be certain that you understand what those ranges/distributions are, what they represent, and how math is being applied.

Precise inputs and outputs

Significance: Solutions that use discrete input values and provide discrete outputs do not accurately represent the reality of cyber risk measurement, and decisions based on these measurements will not be well-informed.

Discussion: Nobody likes uncertainty, but the fact is that every risk analysis is an effort to understand and portray the probability and impact of future loss event scenarios. And because we don't have a perfect understanding of the variables that affect the future, there will always be some amount of uncertainty.

Therefore, one of the most important criteria for realistic and useful risk measurement is that inputs and outputs reflect uncertainty. If we have sparse data for one or more inputs, that fact must be accounted for in the analysis and reflected in the output. Besides improving accuracy, the level of uncertainty in a measurement can be one of the most essential data points for decision-makers because this awareness allows them to:

- If appropriate, be more conservative in their decision-making (e.g., take a safer route or implement greater resiliency).
- Take steps to improve data quality over time if greater precision in the future is desired.

CRQ solutions that fail to faithfully reflect uncertainty in their inputs or outputs should be avoided, at least if you want the results to be realistic and useful for decision-makers.

Use of control standards

Significance: Improperly accounting for control efficacy can invalidate analysis results.

Discussion: Properly accounting for the effect of controls is the most complex dimension of risk measurement. For this reason, despite the value that control frameworks like NIST CSF, NIST 800-53, COBIT, ISO2700x, PCI DSS, etc. provide as a means of identifying control deficiencies, they actually can be dangerous when applied to risk measurement.

The first problem with some of the standards is the fact that the control descriptions have not been defined carefully — at least carefully enough to be clear on how to apply the control within a risk analysis. For example, NIST CSF 1.1 PR.AC-2 is described as "Physical access to assets is managed and protected." This sweeping description defies clear measurement or application within an analysis. Yes, you can take the next step and seek greater clarity within the "Informative References" for this control (e.g., CIS 1, 5, 15, and 16; COBIT 5 DSS05.04, and DSS06.03, etc.) but what you'll almost invariably find is that those references also are loosely described and very often inconsistent with one another to some degree. Absent sufficient clarity about how a control affects the probability or magnitude of loss, the odds are very low of correctly accounting for a control's effect within a risk measurement.

As described earlier, none of the control standards in use today capture the relationships and dependencies between controls. Neither do they capture the varied roles a control can play in affecting the frequency or magnitude of loss for different loss event scenarios. Related to this is the fact that many controls affect more than one aspect of loss event scenarios. In other words, a control (e.g., logging) might function both as a deterrent to a potential bad actor (prevention) and aid in the response effort (response) if a loss

21 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.4544&rep=rep1&type=pdf

event occurs. As a result, a CRQ solution that relies on control standards without a careful (and, unfortunately, very complicated) mapping to logic that accounts for control relationships and scenario relevance will generate unreliable results.

Another common practice used by risk measurement products is to rate control values using ordinal scales (e.g., 1 thru 5, etc.), which are supposed to represent “maturity” or “efficacy.” These ratings are then combined with other variables such as impact and threat levels within a mathematical algorithm to arrive at a risk result. As discussed elsewhere in this document, math on ordinal scales will generate unreliable results.

For the reasons above, correctly accounting for control value in a risk analysis is without question the most challenging aspect of cyber risk measurement. Unfortunately, this is not widely recognized, which means that it is common to see naive methods being used that can completely invalidate analysis results.

Until better controls analysis models are available, reliable risk measurement (whether qualitative or quantitative) still requires human judgment to suss out the nuances in how controls affect various loss event scenarios.

Weighted values

Significance: Weighted values are rarely reliable and can result in unreliable risk measurement.

Discussion: Some cyber risk measurement solutions apply weights to various values in their risk calculations, such as:

- Controls that are believed to be more important than others (e.g., authentication, etc.).
- Impact parameters that are believed to be more important than others (e.g., reputation, etc.).
- Threats communities that are believed to be more dangerous than others (e.g., trusted insiders, etc.).

There are several common problems with weighted values, which can significantly affect the reliability of risk measurement. These problems include:

- Very often, the weighted values are not arrived at with any rigor, which means they are more prone to error.
- Weighted values are often ordinal, which means any math (commonly multiplication) is unlikely to generate reliable results.
- Weighted values are almost always discrete values rather than ranges or distributions, which means they fail to capture the inherent uncertainty in measurements.
- Weighted values are highly sensitive to the scope/context of an analysis. For example, it’s common for preventative controls (e.g., authentication) to be weighted higher than detective controls (e.g., logging). However, in scenarios where the threat is a trusted insider who is supposed to have authenticated access, logging plays a much more important role than authentication. Consequently, an algorithm that weights authentication as the more important control will generate inaccurate results when analyzing this insider type of scenario.

If a risk measurement solution uses weighted values, it is crucial to understand how those values were determined, whether they’re ordinal, how they’re being applied within the analysis, and whether they take into account the numerous scope-related challenges.

“Industry data”

Significance: If improperly applied, even decent industry data can generate unreliable risk measurements.

Discussion: Some CRQ providers claim that the data being applied within their cyber risk analysis is similar to the actuarial data used in mature insurance domains like property and casualty, life, etc. Sometimes they claim to leverage data from tens or hundreds of thousands of insurance claims or other sources. These claims may or may not be valid, but given the quality of data we’ve observed in the industry there are certainly some concerns that should be explored.

- If they leverage insurance claims data you’ll want to understand how they account for forms of loss that usually aren’t covered by insurance.
- The losses from some types of cyber events are highly specific to an organization. For example, the losses that materialize from the theft of intellectual property are going to depend on the value of that particular intellectual property. Consequently, you will want to know how a solution provider extrapolates these kinds of losses from industry data. The same thing is true for losses that materialize from outage events, which is highly dependent upon the affected business process, the recovery capabilities of the organization, their market position, etc.

The bottom line is that when solution providers make this claim, you will want to understand specifics regarding what data they're relying on. Ask for examples of the records they're operating from, and do not let them cherry pick which ones they show you. You also want to understand the assumptions they're making when they apply the data to an analysis. The concern is that in an effort to leverage high volumes of data, they may inappropriately apply data to unassociated variables, or extrapolate too broadly.

The bottom line is that although the availability of data is slowly improving, it's a long way from being anywhere near the quality of standard insurance actuarial data.

"Eliminate guessing"

Significance: Inappropriately applying data and taking shortcuts in risk analytic models will result in unreliable risk measurements.

Discussion: To some extent this red flag is an extension of the previous one. Specifically, some CRQ providers claim their solution is purely data-fed, eliminating the need for expert judgment. However, having data is very different from using data appropriately.

In a perfect world, we would have copious amounts of empirical data that are well-normalized, are aligned to a proven model, and have a reasonable half-life in terms of their usefulness. In fact, there is only a small subset of the cyber risk landscape that today can be reliably analyzed almost entirely using historical data and limited expert judgment, and that's your garden-variety malware infection. Not your sophisticated APT type of malware — just the high-volume stuff.

Because of the noisy, ambiguous, and un-normalized nature of cybersecurity data today, it isn't possible to do cyber risk analysis without involving expert judgment. Consequently, what CRQ solution providers really mean when they claim to eliminate expert judgment is that they're removed the need for you — the user of their solution — to apply expert judgment. They're essentially claiming to have done all of that for you through their algorithms. This means that any errors in their algorithms are going to systemically effect every analysis their solution provides.

It's important to remember that the variety of potential cyber risk loss event scenarios in an organization is huge, and the differences between these scenarios (e.g., vectors, relevant controls, etc.) can be significant but highly nuanced. The more a CRQ automates its analyses, the less it's going to be able to capture these differences and the less likely it is to generate reliable results.

If you're tempted to adopt a solution that claims to eliminate expert judgment, you should dig very deeply into how they fulfill that promise because this is where short-cuts and gross errors occur (some of which are captured in the other red flags in this document).

"Know your organization's risk in a very short time frame"

Significance: Very likely achieved by violating multiple red flags.

Discussion: Some CRQ solution providers claim to be able to provide an enterprise-wide risk analysis in very little time (e.g., hours). Given the previous red flags, it's obvious why this claim should be viewed skeptically. The simple fact is that analyzing overall risk is a non-trivial exercise for any organization with a complex cyber risk landscape (which is the case for most organizations today).

Common shortcuts that underlie efforts to rapidly analyze an organization's overall risk include extrapolating from average values available through various industry data sources, such as:

- The Ponemon loss magnitude report
- Various threat intelligence provider reports, and
- Verizon's DBIR

Certainly, these can be useful data sources, but reliably applying averages from aggregated sources to any particular organization overlooks the fact that details regarding your organization's risk landscape can profoundly affect how well an industry average represents your organization.

Another common shortcut is to rely heavily on over-simplified and typically inaccurate control framework assessments (discussed as a separate red flag).

In order to understand and rely upon any risk analysis, you need to be able to adjust input values to reflect your organization's reality. There's a reason Einstein is quoted as saying, "You should make things as simple as possible, and no simpler." The undisputed fact is that the cyber risk landscape is complex, and over-simplified analyses can easily generate unreliable results.

Proprietary algorithms

Significance: May limit your ability to understand and defend analysis results. Also may be more prone to error.

Discussion: If a CRQ provider touts their proprietary algorithm, insist on a detailed description of how it works (both in plain English and with their mathematical formulas) and ask whether an independent third party has evaluated it. Incidentally, it doesn't matter how many Ph.D.'s worked on an algorithm, as they too can make mistakes. Furthermore, highly complicated risk algorithms can be fragile to changes in analysis scope and the types of scenarios being analyzed.

That said, there is no reason why a proprietary algorithm can't be good. It's merely a question of how well you understand it and your ability to explain to your stakeholders why the results should be trusted. An explanation that boils down to *"It's proprietary and I don't understand how it works, but it was created by Ph.D.'s"* may not be the ideal answer. This is especially true if you find yourself justifying your results to executives, boards, regulators, or others who are skeptical of risk models that can't be explained in plain English.

Note, however, that it isn't just a matter of whether or not equations have errors in them. Every bit as important is the logic underlying the equations, and the assumptions they're operating from (many of which are covered in the Important Questions section above).

Simplistic Aggregation

Significance: Done improperly, this can generate profoundly inaccurate results.

Discussion: Accurately aggregating loss exposure from multiple analyses is a non-trivial problem due to the probability component of an analysis as well as analysis scoping considerations. Unfortunately, some solutions just add up the loss exposures from multiple analyses, which is not going to provide accurate results.

Correct aggregation methods are too involved to describe here, so you should ensure that:

- Risk values are not being added to one another without proper consideration of the probabilities, and
- Loss event scenarios included in aggregation have been very carefully scoped to avoid overlap (see an earlier red flag for more on this).

Spreadsheets

Significance: May introduce analysis integrity and security issues.

Discussion: Spreadsheets can be incredibly useful tools, and there is absolutely nothing inherently wrong with their use in performing CRQ. That said, if a solution provider is using a spreadsheet for their data input, calculations, and reporting, you will want to keep a few things in mind:

- Spreadsheets do not provide secure data storage and tend to find their way into e-mails, mobile media, dark corners of network storage, etc.; so, where and how would your information be secured?
- Spreadsheets are notorious for subtle data corruption and miscalculation in complex analyses. How, and how frequently, is the spreadsheet validated for integrity, and how is version control managed for these spreadsheets?
- Unless steps are taken to secure cells containing the formulas being used, it is possible for users to intentionally or accidentally make alterations that result in unreliable results.

A provider who depends on spreadsheets as their analytic tool may be relatively new to CRQ, or they may not want to get into the software arena. By itself, this doesn't mean they aren't competent or perhaps even very good at performing risk analysis.

Summary

When properly designed and applied, CRQ enables far more cost-effective cyber risk management than can be achieved using other risk-related measurement approaches. It also helps executives to better understand and take into account the significance of cybersecurity concerns in their decision-making.

However, reliably performing CRQ requires:

- Careful and clear scoping of the loss event scenario(s) being measured,
- A well-designed analytic model, and
- Appropriate use of whatever data are available.

It's important to note that these are required regardless of whether measurement is being done qualitatively or quantitatively.

As a relatively new discipline though, it is easy to design CRQ poorly or select a CRQ solution that doesn't generate reliable results. This is particularly true when trying to take shortcuts in order to perform risk measurement at scale. It's essential to keep in mind that unreliable risk measurement done rapidly and at scale (whether quantitative or qualitative), accomplishes nothing more than to systematize poorly informed decision-making.

Hopefully, the information in this paper helped clarify cyber risk measurement methods in general and CRQ in particular, enabling you to ask better questions, recognize potentially problematic solutions, and make better decisions.

There undoubtedly will be those in the industry who disagree with points made in this paper, or that feel this paper is biased in some fashion. Of course, it's always possible that something in this paper is factually incorrect or misses a particularly salient point. With that in mind, public or private discussions are welcome regarding the points made here, in case corrections are warranted or additional clarification is needed.

Addendum

Questions checklist

Category	Question	Notes
Questions related to utility	How do they define the “risks” being measured?	
	What part(s) of the risk landscape does their solution analyze?	
Questions related to data	What data do they use for the frequency of attacks/events?	
	Where does control-related data come from?	
	Where does impact data come from?	
	Is impact data cleanly broken out into categories?	
	What assumptions do they make when applying historical data?	
	How is data uncertainty accounted for?	
	How is fallout from an event (e.g., fines & judgments) analyzed?	
	How are data and/or analyses updated as the risk landscape evolves?	
	What distribution types are they using?	
Questions related to analytics	Does their algorithm perform math on ordinal values?	
	How is control efficacy measured and represented?	
	Does it support what-if analyses?	
	How has their model been validated?	

Category	Question	Notes
Questions related to reporting	Does it break out results into probability and magnitude components?	
	Does it aggregate risk?	
	Is it able to report risk levels by organization unit, line of business, etc.?	
	Does it provide cost-benefit results?	
	Does it report risk level trending over time?	
	Does it enable reporting against risk appetite?	
	Can it export to different formats?	
	How is uncertainty reflected in results?	

Red Flags checklist

Red flag	Yes	Yes, but...	No
"Risks" that aren't risks			
Scope overlap			
Ordinal measurements			
Precise inputs and outputs			
Use of control standards			
Weighted values			
"Industry data"			
"Eliminate guessing"			
"Know the organization's risk in a very short time-frame"			
Proprietary algorithms			
Simplistic algorithms			
Spreadsheets			

- "Yes" indicates that a CRQ solution is subject to this concern and therefore the reliability of its analytic results is uncertain.
- "Yes, but..." indicates that a CRQ solution is subject to this concern but has passed a due diligence examination that allows for confidence in its analytic results. An explanation of the due diligence process and the basis for acceptance should be carefully documented.
- "No" indicates that a CRQ solution is not subject to this concern.

Jack Jones is Chairman of the **FAIR Institute** and creator of Factor Analysis of Information Risk (FAIR), the leading model for quantitative analysis of cyber, technology and operational risk.

Read Jack's book:
Measuring and Managing Information Risk: A FAIR Approach.

The FAIR Institute is an expert, non-profit organization led by information risk officers, CISOs and business executives, created to develop and share standard information risk management practices based on FAIR. Factor Analysis of Information Risk (FAIR) is the only international standard quantitative model for information security and operational risk. FAIR helps organizations quantify and manage risk from the business perspective and enables cost-effective decision-making. To learn more and get involved visit **FAIR Institute.**