GRG Remuneration Insight 136

4 Key Mistakes in Benchmarking to Avoid in 2022

Authors: Denis Godfrey, James Bourchier & Peter Godfrey

INTRODUCTION

Benchmarking KMP remuneration may seem like a simple process, but it can quickly and easily go wrong unless some key sense checks are completed throughout the process and consistently over time. When it goes wrong it can lead to overpayment (excessive cost), loss of talent (disruption), poor performance alignment or shareholder backlash that can lead to unwanted media attention, damage to the reputation of the Board or a spill. In this Remuneration Insight we explore 4 pitfalls to avoid when benchmarking in 2022, including key questions to ask, using the example of a fictitious online electronics retailer, Nogan Limited. While the first may be obvious, which is around developing a benchmarking policy that goes deeper than just "P50" (median, 50th percentile or middle of the sample), COVID-19 has created some bear-traps that need to be avoided when benchmarking in 2022: dealing with volatility in performance rewards and dealing with volatility in market positioning. Lastly, any benchmarking that is done on title matching only is likely to be a "loose fit", so identifying a methodology that can address the role designs and role relationships that are particular to your organisation and no one else's is a key aspect of benchmarking that often requires not only data, but some level of judgement and advice.

AVOID PERFORMANCE IMPACTED BENCHMARKING ERRORS (ESPECIALLY DURING COVID)

The COVID-19 pandemic has created a period of high volatility in market data used for benchmarking, particularly performance-pay. We have seen the full spectrum of reactions, from cancellation of all performance pay in companies hit hard by the pandemic, through to one-off additional bonuses and Board discretion being applied to increase variable rewards in companies that have thrived. What does that mean for benchmarking? Two things: firstly, statutory market data (or private survey data on actual pay for that matter) is showing a much wider range of outcomes than may be typical, and second, overall market data has come down for many samples, due to nil or negative values being reported. This means that raw market data is unlikely to be a good guide for variable remuneration benchmarking at the moment, and experienced analysts and advisors will be making adjustments to market data to determine a sustainable recommendation. At GRG we use our database of over 1,000 companies per year over 20 years to smooth out short term volatility and see through to stable, sustainable variable remuneration practice evidence (i.e. we can use historical relationships/trends analysis to smooth out short term volatility). Alternatively, we can look to "policy" statements given in Remuneration Reports to calculate the intended package for competent performance, rather than the statutory package or "realised" package which is impacted by actual performance outcomes.

This has highlighted a deeper methodological and philosophical issue: should you be benchmarking total packages against statutory market data, maximum/stretch policy data or target data? Given the volatility in statutory/actual pay, it would seem logical to refer to policy benchmark data which is unaffected by actual performance and should therefore be valid and reliable. In this regard there are two key matters to consider:

1. Most companies only report statutory market data, so if you look to policy level data for benchmarking, samples quickly become vanishingly small outside of the ASX 50, however:

a. If you determine to benchmark against stretch/maximum policy data, you will quickly run into the problem of "comparing stretch performance expectations": the stretch or maximum performance concept is perhaps the most volatile and varied of all benchmarking metrics, and is therefore generally a poor candidate for benchmarking. For some companies, stretch is truly blue-sky, while for others, it may be very close to budget. Without being able to establish whether your stretch expectations are close to the typical stretch expectations in the sample, you may be comparing quite different reward structures.

b. If you determine to benchmark against "Target" policy data, you will have to rely on some level of interpretation by analysts, as not all companies use the term "Target", with some using "meets expectations" or "budget" in place of Target. However, a well-trained analyst team will know how to identify or interpolate "Target" from ranges of performance and reward scales, and this level of performance and reward is likely to be much more comparable between companies: few have a target/expectation that is far above or below budget. In this sense, "Target" policy benchmark data is likely to be the most valid, but small samples sizes (as applies to all policy data) may make it unreliable.

2. How are you setting variable remuneration in your own business: is it a bonus for exceeding expectations, at-risk remuneration to be reduced if expectations are not met, or some hybrid of the two? When you sum remuneration components to compare your incumbents' packages to the market, are you including the minimum, target or stretch level of remuneration in that number being benchmarked? GRG generally recommends "Target" or "At Expectations" (challenging but achievable, usually around budget level of performance) since any other outcome should not be expected and is therefore of lesser relevance. Conveniently, this tends to align with the best available market data, even statutory market data, which in normal years is in aggregate (across a sample) around target/budget on average, with only the outliers falling at nil or significantly above budget.

AVOID "THE WRONG P50" FOR FIXED PAY

The majority of ASX listed companies have a remuneration policy of paying Fixed Pay at the P50/middle of the market, which on the face of it seems a responsible but competitive position for doing the bare minimum of the job (i.e. before performance is considered). However, without a clear policy around how market samples will be constructed, this policy means nothing. The P50 could be based on any sample of any number of irrelevant companies, producing an outcome that could actually be P90 or completely outside the relevant range of market practice. While this is an obvious opportunity for issues to arise, developing a policy to govern benchmark formation is nuanced and often overlooked. Many benchmarking exercises fall at the last hurdle because a key stakeholder does not accept the basis of the benchmarking sample and was not consulted. To make matters harder, different stakeholder groups will often have different views regarding how to construct a data sample:

1. **Shareholders, proxy advisors and institutional investors will generally assess remuneration on the basis of market capitalisation, usually breaking up the ASX into buckets of say 25 or 50** (e.g. for Nogan, it may be ASX 150 to 175) and benchmark companies against the P50/median of whatever bucket the company falls into, even if it falls at the top or bottom of the range. For companies at the margins of the bucket, this often leads to benchmarks that are skewed up or down depending on how close the company is to the middle of the bucket. That said, it is often useful to run this benchmark to get a sense of how voting shareholders are likely to view remuneration which may be one of several key considerations in determining appropriate increases.

2. **Executives will often be expecting to be benchmarked against industry peers only, and often, only the larger or global peers that they aspire to compete with for customers and talent**, even if there are only a handful of direct peers. For Nogan, management might be looking to benchmark against Amazon as their biggest competitor for example. This often leads to benchmarks that are small, volatile, and arguably both invalid and unreliable – usually skewed

GRG GODFREY REMUNERATION GROUP

upwards. That said, it is often useful to run an "industry group" of key peers to examine competitor practices on a line-by-line, incumbent-by-incumbent level of detail – that is assuming your data provider can give you constituent data (not available from private surveys, only available from ASX public disclosure benchmarks).

3. **"Data-only" service providers or database houses will often push clients to benchmark on the basis of factors that best suit their data sources: often revenue, job evaluation score or job level**, because their databases tend to be made of unlisted "private survey" company data where it is not possible to obtain market capitalisation values and it keeps clients buying into their valuation methodologies, even though remuneration practices often vary in quantum, structure and design compared to the listed market (invalid benchmarks). That said, revenue is a key metric for many companies, and it is sometimes useful to run a revenue group as a secondary group to inform discussion.

4. **In multinational companies, there is often pressure to build a comparator group that includes Oceanic, American, and major European competitors**. For Nogan, that might mean building a group based on the likes of Argos and Amazon alongside Harvey Norman for example. This immediately produces an invalid sample usually skewed towards much higher American variable remuneration practices. This is because the American market pays very differently in terms of quantum, structure and design, compared to the Australian market (as do the other markets); thus combining data from different markets produces "noise" and volatility in the data that provides no clear insights into the practices of any market. That said, it is often true that international markets are a key consideration, so it may be important to build a secondary or supplementary data sample based on "pure" international markets (one for America, one for the UK for example) such that a clear understanding of how those markets pay, as distinct from one another, can be obtained.

5. **Genuine remuneration advisors should apply best practices in statistical methodology** and try to build tailored data samples around your business such that the measures of central tendency (median/P50 ideally, noting that averages are often skewed), will be highly relevant, reliable and valid, including industry peers to the extent possible. At GRG, for a primary comparator group we generally recommend a sample of at least 20 ASX listed companies, balanced (10 larger and 10 smaller) and within a range of around half to double the target market capitalisation, such that the P50 of sample company values is close to the market value of the client, and such that samples sizes will be robust while limiting the range of companies to those that are of comparable scale. **That said, it is often useful to run at least one secondary or supplementary group from the foregoing list to ensure that multiple views are considered**. In our experience, these are generally discarded in later discussions more often than not, given the strength of a primary statistically valid sample.

AVOID UNSUSTAINABLE BENCHMARKS ARISING FROM COVID RELATED VOLATILITY IN MARKET CAPITALISATION (YOURS AND THE MARKET)

A key challenge for companies benchmarking during the pandemic has been dealing with volatility in share prices which has resulted in significant movement in market positioning (ASX ranking by market capitalisation). Given that market capitalisation and relativity to other companies on the ASX is often a key consideration for benchmarking, decision makers need to be careful about using a benchmark based on a temporary peak or trough in market value or market positioning. While it is tempting to re-benchmark at a peak, particularly as this often occurs in conjunction with increased pressure on talent acquisition and retention (for example in the technology sector where share prices spiked and the war for talent has become intense) – if the share price and market relativity is not sustained through to the next AGM, reactionary increases in executive remuneration may ultimately be viewed as excessive. While it has always been important to use a "forward-looking and sustainable" market capitalisation value for benchmarking to ensure that remuneration levels are appropriate and justifiable long term, the pandemic has made this particularly challenging. We are seeing a couple of typical responses to this tension:

1. Boards requesting that a comparator group be built around a market capitalisation based on a volume weighted average price (VWAP) over a full 12 months, and/or
2. Requesting that all potential comparators be subject to a 12 months VWAP calculation before determining the group.

While at face value this seems a sensible way to smooth out short term volatility, it does not necessarily ensure that the market positioning that results from smoothed market capitalisations is appropriate. While these types of calculations can be used to inform the discussion, it may instead be better to identify companies that have experiencedvolatility very different from your business and exclude them (or use them as comparators with caution) and to focus on what the forward looking expectations are for your own share price rather than assuming that a VWAP of the past is likely to be a good indication of future market positioning – effectively for the Board to agree on a market capitalisation to use for benchmarking purposes even if it is not current or based on a VWAP. While this is a more challenging discussion, it is more likely to avoid anomalous outcomes.

For the Nogan example, which experienced a tech and online retail boom during the height of the pandemic in 2021, the market capitalisation increased around 50% at its height, while its brick-and-mortar peers fell significantly. With the Board now expecting some of those technology and online retail related premiums to fall across the ASX, and for brick-and-mortar peers to do better as COVID restrictions ease, it was decided to use pre-pandemic market capitalisation values (January 2020) to build a comparator group for benchmarking, as the best indicator of market relativities and peer relationships.

AVOID BLINDLY TITLE MATCHING, ESPECIALLY WHEN ROLE IMPACTS ARE SHIFTING

The pandemic has resulted in significant changes in the way many organisations work, which has flowed through to both changes in organisation design, and in role design/scope – for example both technology and HR roles have been critical to getting companies through the pandemic, and it is now difficult to argue that these roles are not classifiable as KMP. The impact on other roles has often been more subtle.

It has always been the case that a simple title-match benchmark is insufficient in that it will fail to address any variances between the role design/scope/impact in your business, and the typical role in the market being benchmarked against, or how the role fits into the overall organisation design in terms of role relativity. The pandemic has only emphasised the need to be able to be sensitive to these issues when benchmarking remuneration. As a result, it is important not to blindly apply a P50 Fixed Pay benchmark for example, but to develop a methodology to identify variances between benchmarks and your own roles, and to be able to make adjustments to market data, either upwards or downwards, to accommodate for these aspects. It is not necessary to undertake full job evaluation or job mapping exercises to achieve this; the Board and CEO typically understand these variances very well and can workshop internal and external role relativities as part of working with a consultant who is able to give advice and recommendations, not just market data.

CONCLUSION

Prior to the pandemic, under a business-as-usual state, many companies have been able to get away with extracting data from a database such as is used to benchmark general employee populations, to do simple role matching for executive roles and arrive at a benchmark to inform remuneration decisions without advice or recommendations. The pandemic has created so much volatility and change that this is now a risky and unsustainable approach – expert advice is likely to be needed to make adjustments to market data, use historical data and apply judgement at every step, from comparator selection, to role relativity analysis and variable remuneration setting, to ensure that practices are appropriate, sustainable and defensible as we move through the pandemic and arrive at an emerging "COVID normal".