Hidden Handicaps of Benchmarking: Impact of the Choice of Industry Classification Taxonomy on Peer Group Based Evaluations

Andrew Banasiewicz
Merrimack College

Benchmarking is one of the most widely used business assessment tools. It is rooted in the use peer group-framed comparisons geared toward objective assessment of company-specific performance, processes, or threat exposures. A critical aspect of benchmarking is peer group selection – here, it is widely believed that the use of standard industry classification offers an unbiased and uniform benchmark-setting basis. More specifically, the notion of 'industry segment' is often implicitly assumed to be definitionally singular, in the sense of being universally defined – that, however, is not the case. There are multiple competing industry definition and classification schemas and cross-taxonomy differences are substantial, which suggest that the choice of industry classification taxonomy can have material impact on the resultant benchmarks, and more importantly, on benchmarking-based evaluations. This article offers an empirical examination of that supposition, built around benchmarking-based assessment of a pharmaceutical company's exposure to securities litigation; the results offer supporting evidence of material impact of the choice of industry classification schema on benchmark values and the resultant decision-making implications.

Keywords: benchmarking, industry classification, shareholder litigation benchmarking, peer group selection

INTRODUCTION

Within the confines of organizational management, it is hard to think of another informational tool that is as widely used as benchmarking, a coarsely defined process used to compare process and performance outcomes with objective standards or practices (Leibfried & McNair, 1997; Seybert, 2006). Given that lack the operational specificity, benchmarking can take on many different forms, but its numerous embodiments can be grouped into two general categories of largely qualitative best practices, and predominantly quantitative indexing (Banasiewicz, 2019). The former encompasses a wide array of processes or procedures that have been shown to be superior to any alternatives, whereas the latter entails the use of objectively derived standards to evaluate the efficacy of performance outcomes, threat exposures, and other types of objectively measurable facets of organizations.

Within the realm of organizational management, benchmarking is ubiqutuous (Banasiewicz, 2019), and its appeal stems from an almost instinctive human tendency to evaluate themselves and others, as originally postulated in the early 1950s by social comparison theory (Festinger, 1954). Further strengthening that argument, recent neuroscientific research suggests that human brain is likely wired for categorization (Branan, 2010), which in turn implies that the instinctive drive toward comparative evaluation can be seen as an expression of basic human sensemaking processes. Given that at their core organizations are human

collectives bound together by the pursuit of common goals (Banasiewicz, 2021), the desire to compare various aspects of organizational functioning to other organizations, typically those seen as peers or high performers, is deeply ingrained in organizational psyche; in a sense, broadly conceived benchmarking can be seen as a natural element of organizational functioning (Banasiewicz, 2019).

A core element of any benchmarking initiative is the identification of relevant and reliable comparison or evaluation benchmarks, or points of reference. Currently, there are no universally accepted criteria for benchmark selection other than the intuitively obvious expectations of relevance and dependability; in applied business benchmarking, using a composite of companies comprising an applicable industry, such as pharmaceutical or automotive, is often considered to offer a maximally unbiased points of reference. However, industry classification itself is not singularly objective because there are numerous competing classification schemas which produce materially different industry structures (Krishnan & Press, 2003; Li et all, 2020). Moreover, even a single classification schema can produce different benchmarking outcomes if its classificatory logic is applied to static vs. dynamic data source (Katselas et al., 2019). All considered, what may be seen as an objective and dependable benchmark may in fact be a point-in-time snapshot of one of several potential depictions of reality. All considered, the widely held belief that the use of established industry classification standards, such as the well-known SIC (Standard Industrial Classification) or GICS (Global Industry Classification Standard) typologies, offers an assurance of an objective, and in a sense 'fixed' reference point, may be unjustified and thus warrants closer investigation.

The ensuing analysis will proceed as follows: First, a high level, descriptive overview of the idea and practice of benchmarking will be undertaken, leading to delineation of clear benchmark-setting criteria. Second, an overview of the most widely used industry classification typologies will be undertaken, from the perspective of using those typologies as benchmark-setting basis. Third, a mini case built around benchmarking of companies' exposure to shareholder litigation will be presented, with the goal of illustrating the impact of industry classification taxonomy on the ultimate risk exposure conclusion.

The Idea and the Practice of Benchmarking

Although embryonic forms of the idea of systematic assessment and evaluation can be traced back to 19th century textile mills (Bogan & English, 1994), the modern conception of that practice is connected to the emergence of formalized quality management practices in the 1950s in the US and Japan, perhaps best illustrated by W. Edwards Deming's Plan-Do-Check-Act cycle used for the control and continuous improvement of business processes[1]. Around that time, in the US, General Electric Company began to experiment with the use of statistical means of evaluating alternative approaches to basic functional activities, while in Japan, Toyota Motor Company began developing its 'kaizen', or continuous improvement manufacturing processes. However, the practice known today as benchmarking is most directly linked to 1980s Xerox Corporation's initiatives geared at stemming its market share slide by systematically tracking activities of its Japanese competitors and making adjustments based on the resultant insights; in fact, the first book dedicated expressly to benchmarking (Camp, 1989) was authored by Xerox's head of benchmarking initiatives.

Though ubiqutuous, benchmarking eludes clear operational specification – it has even been characterized as 'ambiguous, multidimensional and contingent' (Buchanan et al. 2005); in fact, the notion of what, exactly, constitutes benchmarking is seen as so vague that efforts to undertake systematic review of benchmarking focused literature have been deemed impractical, even undesirable (Francis & Holloway, 2007). The rich diversity of benchmarking perspectives is reflected in a number of competing typologies which include internal vs. competitive vs. functional vs. generic (Camp, 1995), results vs. process (Trosa & Williams, 1996), voluntary vs. compulsory (Bowerman, 2002), unilateral vs. cooperative (Elnathan et al., 1996), implicit vs. explicit (Schofield, 1998), and international vs. global vs. external vs. collaborative (CIPFA, 1996). Somewhat hidden within those competing framings of benchmarking is the distinction between performance assessment, which emphasizes objective quantitative measures, and best practices delineation, which relies on largely qualitative comparison of processes or activities. All considered, while nearly universally appealing to organizational decision-makers, lacking consistent conceptual and methodological standards applied benchmarking can be characterized as an informational free-for-all.

It is important to note, however, that the idea of benchmarking is also applied in rigorous scientific research. For example, a recent study focused on the behavior of highway traffic seen as a compressible fluid and expressed through aggregate state variables of flow and density used rigorous benchmarking to assess the efficacy of competing models' performance in real world traffic applications (Mohammadian et al., 2021); another recent study (Albahri et al., 2020) utilized benchmarking for taxonomic evaluation of competing artificial intelligence techniques used in the detection and classification of COVID-19 medical images. Methodologically robust benchmarking approaches have also been proposed to assess the implementation of lean manufacturing principles to healthcare (Reponen et al., 2021) or hotel service quality (Park & Lee, 2021), to support targeted guidance for individual retail stores (Almohri et al., 2018), to assess the overall energy usage (Chen et al., 2018) and energy efficiency of buildings (Kavousian & Rajagopal, 2014) and air conditioning systems (Zhou et al., 2019). Method-wise, scientific benchmarking approaches utilized physical simulations focused on testing specific design parameters, statistical estimation techniques such as regression, cluster, data envelopment, and stochastic frontier analysis, and hybrid methods such as Bayesian analysis or Monte Carlo simulations. However, such methodologically sophisticated benchmarking approaches are rarely used in applied context of organizational management, where simpler compare-and-contrast assessment techniques tend to be preferred (Banasiewicz, 2019).

In a more general sense, whether it is meant to serve as a mean of assessing performance outcomes of business entities or the basis of scientific assessment, the rationale embedded in benchmarking analyses is based on an implicit assumption that assessment benchmarks represent objective norms or standards. And to be sure, there are numerous situations in which that is indeed the case – for instance, the American College of Radiology developed the Breast Imagining Reporting and Data System in the 1980s, which was further solidified by the passage of the Mammography Quality Standards Act by the US Congress in 1992 (Lehman et al., 2017). Still, given the practically innumerable applied and scientific benchmarking contexts there is no reliable way of knowing what proportion of those efforts are tied to such singularly objective and uniform evaluation standards, and what make use of ad hoc evaluation basis. Within the confines of applied benchmarking, anecdotal evidence suggests that the latter might in fact be far more common, as business organizations in particular tend to favor the use of peer groups of their choosing as basis for assessment of key performance outcomes, risk exposure, and other facets of organizational functioning. In that context it is commonly held that using the widely accepted industry classification standards, such as the well-known SIC (Standard Industry Classification) taxonomy, offers a dependable basis for defining a comparative benchmark, but there are reasons – explored in this paper – to believe that it might not be the case.

Benchmark-Setting Criteria

The idea of using benchmarking to evaluate different facets of organizational functioning implies several distinct prerequisites that have to be met by any prospective point of reference, most notably:

1. Relevance, framed here as close adherence to needs and/or interests of users;
2. Objectivity, seen here as impartiality and/or the absence of bias;
3. Reliability, manifesting itself as the capability to deliver comparable results across applications;
4. Validity, or truthfulness;
5. Usability, or ease of administration and interpretation.

Visibly missing from the above enumerated benchmarking-setting criteria are stability and uniqueness related considerations, which are theorized here to be properties of how benchmarks are applied, rather than of benchmarks themselves. More specifically, since benchmarks used in business and similar analyses are ultimately products of dynamic environmental factors, they can only be assumed to be fixed points of reference at a point in time, which implies a systemic lack of stability. In fact, it can be taken to be intuitively obvious that benchmark reliability and validity, and to a lesser degree benchmark relevance and objectivity call for dynamic adaptation. The reasoning behind the absence of uniqueness related benchmarking prerequisites is somewhat more obtuse but can be boiled down to the relativistic[2] character benchmarks'

information foundations, as illustrated by competing industry classification schemas discussed below. Simply put, the same reality can support numerous competing interpretations.

Still, a core part of any benchmarking exercise is the selection of an appropriate evaluation baseline, or a benchmark. Empirical evidence suggests that benchmark selection can have a profound impact on critical organizational decisions – for instance, a large-scale study of pension funds' returns found that benchmark selection was more important in driving returns than investment selection and timing (Broeders & De Haan, 2018); similarly, an investigation of benchmark writing samples concluded that the assessed quality of writing was strongly influenced by the benchmarks chosen to define the evaluation rubric (Popp et all, 2003). In the applied realm of organizational management, it is widely held that industry index, typically expressed as an average of companies comprising a particular industry, offers an objective, and in a sense fixed (at a point in time) evaluation point of reference. However, what constitutes a given industry is tied to the underlying classification schema, the choice of which can result in materially different industry benchmark (Bhojraj et al., 2003; Krishnan & Press, 2003; Li et al., 2020). And there is a plethora of methodologically robust, but ultimately typologically incommensurate schemas.

INDUSTRY CLASSIFICATION TAXONOMIES

Currently, there are numerous industry classification standards that have been developed over the past several decades by a mix of national governments, transnational bodies, and private organizations. National government-drafted industry classification taxonomies include the Standard Industrial Classification (the oldest industry classification taxonomy and the source of the ubiqutuous SIC codes) created by the United States government, the North American Industry Classification System developed by the United States, Canadian and Mexican governments, United Kingdom Standard Industrial Classification of Economic Activities drafted by the United Kingdom government, Swedish Standard Industrial Classification developed by the government of Sweden, Australian and New Zealand Standard Industrial Classification created by governments of Australia and New Zealand, and the European Union-developed Statistical Classification of Economic Activities in the European community. Transnational bodies-conceived taxonomies include International Standard Industrial Classification of All Economic Activities and United Nations Standard Products and Services Code, both created by the United Nations. There are numerous private interests-developed taxonomies, a group which is perhaps best exemplified by Standard & Poor's and MSCI co-developed Global Industry Classification Standard, and FTSE-developed Industry Classification Benchmark. Table 1 offers a summary of the ten (10) best-known industry classification taxonomies, looked at from the perspective of applied, US-based users.

TABLE 1
THE BEST-KNOWN INDUSTRY CLASSIFICATION TAXONOMIES

Industry Classification Scheme	Year Introduced	Orientation	Geographic Scope	Classification Units	Hierarchy Levels	Update
Standard Industrial Classification (SIC)	1937	Production and market	United States	Establishments	4	No longer updated (since 1987)
International Standard Industrial Classification (ISIC)	1949	Production	Global	Establishments	4	Last update: 1989
North American Industry Classification System (NAICS)	1997	Production	North America	Establishments	5	Every 5 years
Global Industry Classification Standard (GICS)	1999	Market	Global (120 countries)	Companies; Securities	4	Annual reviews
Morningstar Category Global Equity Classification Structure	2000	Market	Global	Companies; Securities	4	Ad hoc
Industry Classification Benchmark (ICB)	2001	Market	Global (75 countries)	Companies; Securities	4	Biannual reviews
FactSet Revere Business and Industry Classification System (RBICS)	2002	Market	Global (78 countries)	Companies; Securities; Business lines	6	Annual reviews
Thomson Reuters Business Classification (TRBC)	2004	Market	Global (130 countries)	Companies; Securities	5	Ad hoc
Bloomberg Industry Classification System (BICS)	2011	Market	Global	Companies; Securities; Business lines	7	Annual reviews
Sustainable Industry Classification System (SICS)	2013	Impact	United States	Companies	3	Unknown

The competing taxonomies summarized in Table 1 differ, most notably in terms of their orientation, which tends to be either production-, i.e., emphasizing process similarities, or market-, i.e., emphasizing demand characteristics, centric[3], geographic scope, classification units (companies, establishments, business lines, securities), and hierarchy levels. Moreover, there is also a considerable amount of cross-taxonomy update cycle variability, with some, like the SIC codes no longer being updated at all, while others, such GICS, receiving annual updates, and still others, such as TRBC being updated on ad hoc basis.

As can be surmised from the cross-taxonomy differences highlighted in Table 1, those competing descriptions of the structure of industrial activities can generate conflicting descriptions of product market competition and firm characteristics across product market competition levels (Li et al., 2020). In fact, different classification systems are seldom consistent for a given firm – for instance, a study by Krishnan & Press (2003) found that mapping four-digit SIC codes to five- or six-digit NAICS (which was introduced in 1997 expressly to replace the SIC structure dating back to the 1930s) produced only 41.9% agreement; in a similar study, Bhojraj et al. (2003) found only 56% agreement between GICS and SIC classifications[4]. It thus follows that the choice of industry classification taxonomy will likely have direct and material impact on benchmark-setting, and by extension on subsequent benchmarking analyses and conclusions, which warrants closer investigation.

One of the best contexts for examining the nature and the extent of potential taxonomy choice-related benchmarking dependance is offered by the ever-present threat of securities litigation, which encompass allegations (by shareholders or regulators) of failure to fully discharge managerial duties on the part of directors and officers of, typically publicly traded, business enterprises. Broadly known as 'executive risk', the threat of shareholder litigation is often assessed in the comparative context of industry peer group-based analysis, discussed in more detail next.

Shareholder Litigation, Industry Classification, and Exposure Benchmarking

One of the most visible obligations of managers of companies with stocks traded on public exchanges, i.e., public companies, is timely, accurate and complete disclosure of pertinent performance related information (Kross and Suk, 2012; Morris et al., 2012); in the US, those expectations are shaped by federal laws, most notably the US Securities Act of 1933 and the US Securities Exchange Acts of 1934 (Palmiter, 2009; Pickering, 1968; Sametz, 1991). Allegations of failure to fulfill those requirements is the essence of

securities litigation, though it is worth noting that the definition of 'disclosure' includes both written, such as formal annual financial filings (e.g., US SEC Form 10-K) as well as informal verbal communications, such as comments made during analyst calls (Kross and Suk, 2012; Lees, 1981). Moreover, the rights of investors in that regard are absolute, which means that no distinction is made between intentional and unintended errors, omissions, or misstatements (Morris et al., 2012). In legal terms, any company performance-related disclosure error or omission can be construed as a violation of securities laws, even if no discernible intent to deceive was alleged (Chen, 2010; Emanuel, 2005). Lastly, under the US law, shareholders (or other qualified parties) seeking legal relief can sue for monetary damages, which they typically do as group known as 'class' (Bebchuk and Fried, 2004; Chung and Wynn, 2008; Harris, 2008), hence those suits are commonly known as shareholder class actions, or SCAs (Barabanov et al, 2008; Francis et al., 1994; Johnson and Clearfield, 2006).

The threat posed by securities litigation is twofold: 1. reputation-damaging negative publicity, and 2. potentially substantial monetary losses[5]. Given the significant economic and reputational risk associated with those events, virtually all public, US-operating companies, which encompasses all entities doing business in the US without regard to where they are domiciled herein (a business entity does not have to be based in the United States to be subject to US securities laws), purchase a specific type of liability protection commonly known as directors' and officers', or D&O for short, insurance. A critical part of determining the most appropriate D&O coverage protection is an objective assessment of securities litigation threat exposure[6], which is commonly expressed in terms of two independent dimensions of likelihood of becoming a target of securities litigation, and severity or the most likely cost (Banasiewicz, 2015). And while there are numerous approaches to estimating those two facets of executive risk[7], peer benchmarking almost always plays an important contributing role.

Mini Case: Benchmarking SCA Exposure of a Pharmaceutical Company

Occurrence-wise, securities litigation is a comparatively infrequent event (Banasiewicz, 2015). More specifically, over the past several years there have been on average between 300 and 400 securities class action lawsuits filed in the US courts annually (Securities Class Action Clearinghouse, 2021), which may seem high, but in fact is relatively modest considering that just the two major US stock exchanges – the New York Stock Exchange and NASDAQ – list about 2,800 and 3,300 stocks, respectively, in addition to roughly 12,000 stocks that are traded over-the-counter (i.e., directly between two parties, without the supervision of an exchange). At the same time, securities litigation can be very costly, not just in terms of monetary damages, but also legal defense costs (see footnote 5). In the vast majority of cases that survive the initial legal discovery (gathering and review of evidence to inform the sufficiency determination), monetary damages take the form of out-of-court settlements[8], the magnitude of which – as summarized in Figure 2 – varies widely across industry sectors (mean and median values), as well as within each sector (standard deviation). Recognizing the inherent difficulty of valid and reliable SCA exposure estimation coupled with the importance of robust risk assessment (Chung & Wynn, 2008; Francis et al., 1994), stock companies tend to utilize multi-pronged risk estimation, typically built around the informational core of peer benchmarking.

TABLE 2
SCA SETTLEMENT VARIABILITY ACROSS GICS-DEFINED INDUSTRY SECTORS

GICS Sector	Mean	Median	Std. Deviation
Communication Services	$ 33,027,317	$ 9,850,000	$ 79,510,022
Consumer Discretionary	$ 25,847,823	$ 5,775,000	$ 57,333,737
Consumer Staples	$ 42,960,176	$ 14,500,000	$ 81,026,686
Energy	$ 48,207,829	$ 9,500,000	$ 98,427,567
Financials	$ 63,471,753	$ 15,600,000	$ 113,308,689
Health Care	$ 40,571,754	$ 8,500,000	$ 93,339,946
Industrials	$ 26,047,647	$ 6,750,000	$ 66,796,385
Information Technology	$ 27,760,999	$ 6,500,000	$ 69,729,532
Materials	$ 20,696,456	$ 6,250,000	$ 50,054,063
Real Estate	$ 38,905,694	$ 9,925,000	$ 76,380,321
Utilities	$ 52,653,826	$ 14,275,000	$ 86,319,887

Central to peer benchmarking is the definition of what constitutes an appropriate peer group. Within the confines of risk assessment, the most commonly used peer group framing makes use of an objectively defined industry segmentation, which divides the universe of business entities into a set of mutually exclusive and collectively exhaustive groupings. Using that general logic, a pharmaceutical company's peers would be all other pharmaceutical companies, financial services firm's peers would be all other financial services firms, etc. Recalling the earlier discussed five (5) benchmark-setting criteria of relevance, objectivity, validity, reliability and usability, reliance on objectively derived industry groupings appears to satisfy those criteria, but that is only if one assumes the existence of a single, universally accepted industry classification schema. However, as discussed earlier and as summarized in Table 1, that is not a valid assumption. Not only are there numerous competing industry classification schemas, but those distinct taxonomies produce significantly different industrial structures which ultimately gives rise to taxonomy-choice-laden benchmarks. When considered in the context of estimation of exposure to securities litigation, the dependence of exposure benchmarks on the choice of industry classification schema undermines the very idea of exposure benchmarking, which implicitly but strongly assumes uniqueness.

Practical implications of taxonomy-choice-laden benchmarking dependency are examined next using the case of a real-life pharmaceutical company, referred to as PharmaCo here. The ensuing analyses utilized data sourced from Erudite Analytics' SCA Tracker®, a proprietary database tracking filings and dispositions (mostly notably, settlements) of securities class actions from 1996 onward; the SCA Tracker® encompasses 8,857 companies and contains details of 2,323 individual securities litigation related settlements. The impact of the choice industry taxonomy is assessed by comparing results emanating from three distinct peer defining perspectives: NAICS, GICS, and SEC. The NAICS and GICS schemas, are the two best-known actively updated formal industry classification taxonomies, whereas the informal SEC grouping represents the securities laws enforcement agency's view of the industrial structure. The 8,857 companies comprising the SCA Tracker® were divided into taxonomy-based industry segments, using criteria delineated by individual schemas; to enhance the clarity of cross-taxonomy comparisons, non-specific groupings, such as 'unclassified' or 'other' were excluded from the analysis (exclusions had no impact on the number of companies in industry segments of interest). Table 3 offers a summary of the resultant taxonomy-specific industry structure.

TABLE 3
COMPETING FRAMINGS OF A PHARMACEUTICAL COMPANY'S INDUSTRY PEER GROUP

North American Industry Classification System - NAICS	Entity Count
Accommodation & Food Services	42
Admi Support & Waste Management	59
Agriculture & Forestry	58
Arts, Entertainment, Recreation	156
Construction	120
Educational Services	103
Finance & Insurance	324
Health Care	135
Information	343
Manufacturing	**3,737**
Mining, Oil & Gas	760
Other Services	67
Professional, Scientific & Technical Services	452
Real Estate & Leasing	395
Retail Trade	396
Transportation & Warehousing	320
Utilities	140
Wholesale Trade	199

Global Industry Classification Standard - GICS	Entity Count
Communication Services	401
Consumer Discretionary	1,013
Consumer Staples	370
Energy	578
Financials	318
Health Care	**1,324**
Industrials	1,273
Information Technology	1,218
Materials	721
Real Estate	450
Utilities	148

US Securities & Exchange Commission - SEC	Entity Count
Energy & Transportation	1,188
Financial Services	1,121
Life Sciences	**1,649**
Manufacturing	1,382
Real Estate & Construction	842
Technology	1,163
Trade & Services	1,600

There are two immediately visible differences: 1. materially different structures, and 2. distinctly different peer group framings. Starting with the former, the NAICS perspective yields the largest number of industry groupings and the greatest industry size (i.e., count) variability: the Accommodation & Food Services segment encompasses just 42 entities, whereas the Manufacturing sector accounts for a whopping 3,737 entities. On the other hand, the informal SEC taxonomy yields the lowest number of industry groupings and the least amount of entity count variability. More importantly, each of the three grouping approaches gives rise to a different reference point: When considered from the perspective of NAICS PharmaCo is a Manufacturing company, GICS classifies it as a Health Care company, and SEC sees it as a Life Sciences entity. And while at first such overt disagreement might be suggestive of classificatory problems, upon closer examination it becomes clear that all three designations can be deemed appropriate. PharmaCo is one of the major drug developers used in health care thus it is both a manufacturing and a health care company, and in a broader sense can also be seen as a part of the larger life sciences ecosystem. Still, each set of peers – Manufacturing vs. Health Care vs. Life Sciences – is made up of a somewhat different mix of companies, which in turn gives rise to the possibility of material differences in likelihood and severity benchmarks, ultimately precipitating potentially different SCA exposure conclusions. Consider tables 4a and 4b.

TABLE 4A
BENCHMARKING SCA LIKELIHOOD

Classification Schema	Company Count	SCA Count	Benchmark Frequencies
NAICS			
Manufacturing	3,737	204	5.5%
GICS			
Health Care	1,324	93	7.0%
SEC			
Life Sciences	1,649	101	6.1%

The range of benchmark frequencies summarized in Table 4a is noticeable, but not dramatic. A likely potential explanation of the moderate spread can be seen as a manifestation of two well-known phenomena of the law of large numbers and regression toward the mean. Each of the three industry groupings – Manufacturing, Health Care, and Life Sciences – can be seen as a large sample (the smallest segment count is 1,324 companies) drawn from the population of all companies used in classification, and according to the logic of the law of large numbers, that sample's average is expected to approach the true population mean. That drift toward the all-companies average, i.e., regression toward the mean, is further compounded by the longitudinal spread of the SCA tracking data, which goes back to 1996.

TABLE 4B
BENCHMARKING SCA SEVERITY

Classification Schema	Settlements		95% Confidence Intervals	
	Median	Mean	Lower	Upper
NAICS				
Manufacturing	$ 6,950,000	$ 30,765,396	$ 21,247,705	$ 40,283,087
GICS				
Health Care	$ 8,500,000	$ 51,339,099	$ 35,251,657	$ 67,426,541
SEC				
Life Sciences	$ 8,000,000	$ 40,506,971	$ 27,687,055	$ 53,326,886

Although the severity aspect of SCA exposure benchmarking is subject to the same statistical forces, there are nonetheless far more pronounced differences among the three industry classification typologies, as summarized in Table 4b, which points toward pervasive taxonomy choice related skew. When seen as a part of the Manufacturing segment, PharmaCo's cost related industry benchmarks are considerably lower than when it is considered to be a part of the Health Care industrial segment. As a result, depending on which one of the three industry taxonomy framed perspectives is used, key risk management decisions, most notably the purchase of applicable liability insurance, would be materially different, as using the Manufacturing segment as the point of reference suggest much lower economic risk that using Health Care to benchmark PharmaCo's SCA exposure.

The three perspectives captured in tables 4a and 4b are just that – three different interpretations of the interpretation-laden reality of SCA exposure. PharmaCo is typical of large industrial enterprises, many – perhaps even most – of which can be seen through multiple lenses, i.e., can be classified into manifestly different but equally valid industry segments. As a major pharmaceutical company, PharmaCo can be seen as a health care, manufacturing, or life sciences entity, and in each case the resulting grouping is a mix of highly similar (i.e., other pharmaceutical companies) and largely dissimilar companies: When looked at as

a health care entity, it is lumped together with hospitals and other health care providers; when seen as a manufacturing firm it is mixed-in with makers of all manner of industrial and household goods, and lastly, when considered a part of the life sciences sector, it's sector membership-defined peer group includes medical device makers, biotechnology and nutraceutical firms, and food processing companies. Here, the information theory framing of communication theory (Shannon, 2011; Murphy, 1991) presents a helpful parallel: When considered from the perspective of benchmarking, industry groupings are mixtures of apposite (signal) and inapposite (noise) entities, from which it follows that the choice of industry classification taxonomy should be informed by the ratio of apposite-to-inapposite company assignments. In general, the higher the proportion of 'alike' companies, the higher the reliability of industry group-based benchmarks.

CONCLUSIONS

The appeal of benchmarking cannot be overstated – it is one of the most widely used decision-aiding informational tools, primarily because it conveys a sense of evaluation objectivity. However, as evidenced by the results of analyses outlined above, the perceived objectivity of peer benchmarking, in which the definition of 'peer group' is framed by industry segment membership, is rooted in an invalid assumption of definitional singularity of industry groupings. When cross-industry classification taxonomy differences are taken into account multiple benchmark values emerge, potentially supporting different benchmarking-based evaluations and managerial choices.

The analysis outlined here was focused on the use of benchmarking to assess companies' exposure of securities litigation, which is but one aspect of enterprise risk management, and which in turn is but one aspect of organizational management. However, even though the conclusions of this research are framed in a relatively narrow context of a particular aspect of organizational functioning, the focal point of the analysis is on expressly differentiating between an 'objective' and a 'universal' frame of reference. Competing industry definition and classification taxonomies, all of which provide objective means of peer group framing, are likely to spawn materially different benchmarks and benchmarking-derived conclusions because there is no single, i.e., universal, industry structure. Whether the focus of benchmarking is to use an industry composite as an objective basis for evaluation of exposure to a particular type of risk or for assessment of performance, the choice of taxonomy used to create industry segments will likely have a material impact on subsequent evaluations and conclusions.

ENDNOTES

1. The term 'benchmark' can be traced back to the emergence of firearms in the 19th century, and more specifically, the process that was used to assess the accuracy of firearms: rifles were fixed in a bench to make it possible to fire several identical shots at a target, and the spread in resultant marks was used as the basis for accuracy assessment.
2. Used here to refer to a philosophical perspective that holds that there is no single, universally true set of standards, and typically contrasted with positivism/objectivism, which holds that there is indeed such a single set of standards.
3. The Sustainable Industry Classification System introduced in 2013 uses impact as classification basis, but thus far it is the only taxonomy that is not built around either production or market focus.
4. It is worth noting that even though the SIC classification schema has been effectively retired in the late 1990s, the SIC codes are still widely used in practice.
5. The median securities litigation settlement value is about $8.5 million, but the top 10 largest settlements all exceed $1 billion, with the largest securities class action settlement topping the $7.42 billion mark; those values do not include legal defense costs, which can run as high as about 40% of settlements.
6. Given that insurance premiums represent expense with no residual value, over-insuring can create a drag on company's earnings, while under-insuring heightens the possibility of inadequate risk protection.
7. For more details, see Banasiewicz, A. (2013), Risk Profiling of Organizations, 2nd ed., Erudite Analytics.

8. Since 1996 (the beginning of the 'modern era' of securities litigation which followed passage of the Private Securities Litigation Reform Act) there have been over 6,000 individual securities lawsuits filed in the US; of those, fewer than 10 were adjudicated by a jury trial (vs. over 2,300 that were settled out of court).

REFERENCES

Albahri, O.S., Zaidan, A.A., Albahri, A.S., Zaidan, B.B., Abdulkareem, K.H., Al-qaysi, Z.T., . . . Rashid, N.A. (2020). Systematic review of artificial intelligence techniques in the detection and classification of COVID-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects. *Journal of Infection and Public Health*, *13*(10), 1381–1396.

Almohri, H., Chinnam, R.B., & Colosimo, M. (2019). Data-driven analytics for benchmarking and optimizing the performance of automotive dealerships. *International Journal of Production Economics*, *213*, 69–80.

Banasiewicz, A. (2015). The Ecosystem of Executive Threats: A Conceptual Overview. *Risk Management*, *17*(2), 109–143.

Banasiewicz, A. (2019). *Evidence-Based Decision-Making: How to Leverage Available Data and Avoid Cognitive Biases.* New York: Routledge.

Banasiewicz, A. (2021). *Organizational Learning in the Age of Data.* Cham: Springer.

Barabanov, S.S., OnemOzocak, H.J.T., & Walker, T.J. (2008). Institutional investors and shareholder litigation. *Financial Management*, *37*(2), 227–250.

Bebchuk, L.A., & Fried, J.M. (2004). *Pay Without Performance: The Unfulfilled Promise of Executive Compensation.* Cambridge: Harvard University Press.

Bhojraj, S., Lee, C.M., & Oler, D.K. (2003). What's my line? A comparison of industry classification schemes for capital market research. *Journal of Accounting Research*, *41*, 745–773.

Bogan, C.E., & English, M.J. (1994). *Benchmarking for Best Practices: Winning through Innovative Adaptation.* New York: McGraw-Hill.

Branan, N. (2010, January). Are Our Brains Wired for Categorization? *Scientific American Mind*, pp. 7–11. doi: 10.1038/scientificamericanmind0110-11

Broeders, D., & De Haan, L. (2018). *Benchmark selection and performance* (603 ed.). Amsterdam: De Nederlandsche Bank. DNB Working Paper No. 603. Retrieved from https://www.dnb.nl/en/binaries/Working%20Paper%20No.%20603_tcm47-378103.pd

Buchanan, D., Fitzgerald, L., Ketley, D., Gollop, R., Jones, J.L., Saint Lamont, S., . . . Whitby, E. (2005). No going back: A review of the literature on sustaining organizational change. *International Journal of Management Reviews*, *7*, 189–205.

Camp, R.C. (1995). *Business Process Benchmarking: Finding and Implementing Best Practices*. Milwaukee: ASQC Quality Press.

Chen, S. (2010). Bolstering unethical leaders: The role of the media, financial analysts and shareholders. *Journal of Public Affairs*, *10*(3), 200–215.

Chen, Y., Tan, H., & Berardi, U. (2018). A data-driven approach for building energy benchmarking using the Lorenz curve. *Energy & Buildings*, *169*, 319–331.

Chung, H.H., & Wynn, J.P. (2008). Managerial liability coverage and earnings conservatism. *Journal of Accounting and Economics*, *46*(1), 135–153.

CIPFA. (1996). *Benchmarking to Improve Performance.* London: Chartered Institute of Public Finance and Accountancy.

Elnathan, D., Lin, T.W., & Young, S.M. (1996). Benchmarking and management accounting: A framework for research. *Journal of Management Accounting Research*, *8*, 37–54.

Emanuel, S. (2005). *Corporations* (3rd ed.). New York: Aspen Publishers.

Festinger, L. (1954). A Theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Francis, G., & Holloway, J. (2007). What have we learned? Themes from the literature on best-practice benchmarking. *International Journal of Management Reviews*, *9*(3), 171–189.

Francis, J., Philbrick, D., & Schipper, K. (1994). Shareholder litigation and corporate disclosure. *Journal of Accounting Research*, *32*(2), 137–164.

Harris, J.D. (2008). Financial misrepresentation: Antecedents and performance effects. *Business & Society*, *47*(3), 390–401.

Johnson, K., & Clearfield, A. (2006). Improving governance by joint shareholder action. *Pensions & Investments*, *34*(5), 12.

Katselas, D., Sidhu, B.K., & Yu, C. (2019). Know your industry: The implications of using static GICS classifications in financial research. *Accounting & Finance*, *59*(2), 1131–1162.

Kavousian, A., & Rajagopal, R. (2014). Data-driven benchmarking of building energy efficiency utilizing statistical frontier models. *Journal of Computing in Civil Engineering*, *28*(1), 79–88.

Krishnan, J., & Press, E. (2003). The North American industry classification system and its implications for accounting research. *Contemporary Accounting Research*, *20*(4), 685–717.

Kross, W.J., & Suk, I. (2012). Does regulation FD work? Evidence from analysts' reliance on public disclosure. *Journal of Accounting and Economics*, *53*(1–2), 225–248.

Lees, F.A. (1981). *Public Disclosure of Corporate Earnings Forecasts.* New York: Conference Board.

Lehman, C.D., Arao, R.F., Sprague, B.L., Lee, J.M., Buist, D.M., Kerlikowske, K., . . . Miglioretti, D.L. (2017). National performance benchmarks for modern screening digital mammography: Update from the Breast Cancer Surveillance Consortium. *Radiology*, *283*(1), 49–58.

Leibfried, K.H.J., & McNair, C.J. (1997). *Benchmarking: A Tool for Continuous Improvement.* New York: Harper Business.

Li, S., Liu, Q., & Refalo, J. (2020). Industry classification, product market competition, and firm characteristics. *Finance Research Letters*, *36*, 1–6.

Mohammadian, S., Zheng, Z., Haque, Md. M., & Bhaskar, A. (2021). Performance of continuum models for real world traffic flows: Comprehensive benchmarking. *Transportation Research Part B*, *147*, 132–167.

Morris, J., Grippo, F., & Barsky, N. (2012). A new era of accountability? *Strategic Finance*, *93*(11), 42–45.

Murphy, M.A. (1991). The mathematical theory of communication. *Communication Research*, *18*(6), 825–833.

Palmiter, A.R. (2009). *Corporations* (6th ed.). Philadelphia: Wolters Kluwer Law & Business.

Park, J., & Lee, B.K. (2021). An opinion-driven decision-support framework for benchmarking hotel service. *Omega*, *103*, 1–20.

Pickering, M.A. (1968). The company as a separate legal entity. *The Modern Law Review*, *31*(5), 481–511.

Popp, S.E.O., Ryan, J.M., Thompson, M.S., & Behrens, J.T. (2003). *Operationalizing the Rubric: The Effect of Benchmark Selection on the Assessed Quality of Writing.* Washington DC: American Educational Research Organization.

Reponen, E., Rundall, T.G., Shortell, S.M., Blodgett, J.C., Juarez, A., Jokela, R., . . . Torkki, P. (2021). Benchmarking outcomes on multiple contextual levels in lean healthcare: A systematic review, development of a conceptual framework, and a research agenda. *BMC Health Services Research*, *21*(1), 1–18.

Sametz, A.W. (1991). *The Battle for Corporate Control: Shareholder Rights, Stakeholder Interests, and Managerial Responsibilities.* New York: Irwin.

Schofield, A. (Ed.). (1998). *Benchmarking in Higher Education: An International Review.* New York: CHEMS and UNESCO.

Securities Class Action Clearinghouse. (2021). Retrieved October 6, 2021, from https://securities.stanford.edu

Seybert, J.A. (2006). *Benchmarking: An essential Tool for Assessment, Improvement, and Accountability.* San Francisco: Jossey-Bass.

Shannon, C.E. (2011). A mathematical theory of communication. *Mobile Computing and Communications Review*, *5*(1), 1–55. (originally published in 1948 in *Bell System Technical Journal*).

Trosa, S., & Williams, S. (1996). Benchmarking in public sector performance management. Performance Measurement in Government. *OECD Occasional Papers*, *9*.

Zhou, Y., Lork, C., Li, W-T., Yuen, C., & Keow, Y. M. (2019). Benchmarking air-conditioning energy performance of residential rooms based on regression and clustering techniques. *Applied Energy*, *253*, 1–16.