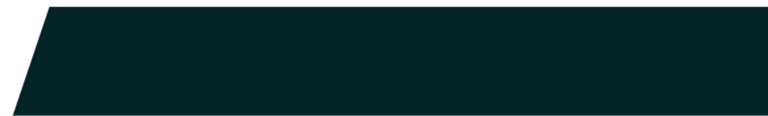


Racial and Gender Diversity’s Effect on Corporate Performance

Executive Summary:

- Many corporate diversity measures, such as boardroom race and gender quotas, are based on a series of McKinsey reports linking executive diversity to corporate outperformance. Those reports are deeply flawed.
- Even if McKinsey’s asserted correlation were true, it does not provide any evidence that racial and gender executive diversity cause outperformance and it only considers one very narrow measure of financial success.
- The correlation the reports find is the product of multiple kinds of cherry-picking.
- It is more likely that financial success causes corporations to adopt diversity measures than the reverse.
- Career opportunities should be distributed according to merit, to whomever would use them to make shareholders the most profit.

Introduction

In theory, diversity of thought strengthens organizations by preventing groupthink; in practice, the business world's pursuit of diversity has become skin-deep. Goldman Sachs recently implemented quotas requiring its portfolio companies and IPOs to have at least one woman and one racial minority on their board of directors.[1] A 2020 Nasdaq Stock Market rule requires listed companies to meet similar board diversity quotas or justify why they have not.[2] To support such measures, financial leaders often claim that racial and gender diversity among a company's top ranks improves corporate performance, though academic studies often disagree.[3] After the Supreme Court's decision striking down university affirmative action, courts will likely enforce Title VII's prohibition on its corporate equivalent, but this paper concerns itself only with evaluating the evidence suggesting these diversity measures increase profit. It finds that evidence deficient.

These quotas, along with a host of other corporate diversity measures, are largely founded upon a series of reports from management consulting firm McKinsey & Company. "Why diversity matters" argued that companies with greater executive racial and gender diversity were more likely to outperform financially. Two follow-ups claimed that the link between diversity and outperformance was becoming even stronger. The most recent, "Diversity wins," illustrates McKinsey's argument by showing how likely the top quartile of companies by diversity were to outperform their industry's average earnings margin relative to the bottom quartile:[4]

---

[1] Ross Kerber, "Goldman Sachs Ups Diversity Targets as Demographic Data Improves." *Reuters*, 2 Dec. 2021, www.reuters.com/markets/us/goldman-sachs-ups-diversity-targets-demographic-data-improves-2021-12-02/.

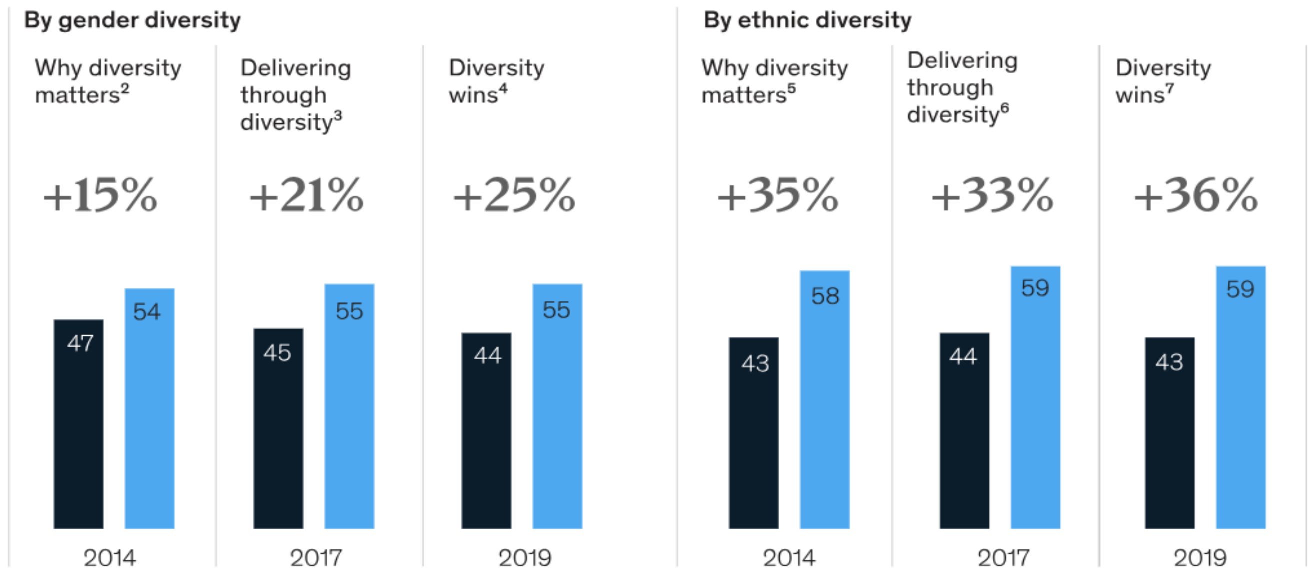
[2] Brian V. Breheny et al, "SEC Approves Nasdaq Board Diversity Listing Standards." *Skadden, Arps, Slate, Meagher & Flom LLP*, www.skadden.com/insights/publications/2021/09/quarterly-insights/sec-approves-nasdaq-board-diversity-listing.

[3] Jesse M. Fried, "Will Nasdaq's Diversity Rules Harm Investors?" *European Corporate Governance Institute*, March 31, 2021, Law Working Paper No. 579/2021, available at SSRN: https://ssrn.com/abstract=3812642 or http://dx.doi.org/10.2139/ssrn.3812642. For an influential study finding that female board members have a negative effect on share value, see Renée B. Adams and Daniel Ferreira, "Women in the Boardroom and Their Impact on Governance and Performance." *Journal of Financial Economics*, Vol. 94 No. 2, November 2009, 291-309, https://doi.org/10.1016/j.jfineco.2008.10.007.

[4] Sundiatu Dixon-Fyle et al, "Diversity Wins: How Inclusion Matters." *McKinsey & Company*, 19 May 2020, www.mckinsey.com/featured-insights/diversity-and-inclusion/diversity-wins-how-inclusion-matters.

The business case for diversity in executive teams remains strong.

Likelihood of financial outperformance,[1] % ■ Bottom quartile ■ Top quartile



[1]Likelihood of financial outperformance vs the national industry median; p-value <0.05, except 2014 data where p-value <0.1. [2]n = 383; Latin America, UK, and US; earnings before interest and taxes (EBIT) margin 2010–13. [3]n = 991; Australia, Brazil, France, Germany, India, Japan, Mexico, Nigeria, Singapore, South Africa, UK, and US ; EBIT margin 2011–15. [4]n = 1,039; 2017 companies for which gender data available in 2019, plus Denmark, Norway, and Sweden; EBIT margin 2014–18. [5]n = 364; Latin America, UK, and US; EBIT margin 2010–13. [6]n = 589; Brazil, Mexico, Singapore, South Africa, UK, and US; EBIT margin 2011–15. [7]n = 533; Brazil, Mexico, Nigeria, Singapore, South Africa, UK, and US, where ethnicity data available in 2019; EBIT margin 2014–18.
Source: Diversity Wins data set

McKinsey
& Company

McKinsey's original 2014 report claims top quartile companies by executive gender diversity were 4% likelier than chance to have an EBIT[5] margin above their national industry average and bottom quartile ones were 3% less likely than chance to outperform the average. This claim is much narrower than the headline "diverse companies perform better," so it is worth breaking down what it means.

When a factor is irrelevant to people's performance on a task, we should expect people with more and less of that factor to perform the same at that task; both groups should perform at the average. For instance, we expect height to give no advantage on the

[5] Earnings before interest and taxes.

SAT, so we would expect tall and short people to get average scores, overall. That means half of short people should get above-average scores and half should get below-average ones; likewise for tall people. If tall people got above-average scores 60% of the time and short ones got them only 40% of the time, that would be a little surprising—it would mean being tall is correlated with a greater chance of a high SAT score.

McKinsey is applying this reasoning to gender diversity and earnings margin. If gender diversity had no effect on earnings margin, one would expect companies in both the top and bottom quartiles of gender diversity to have a 50-50 chance of outperforming their industry's average. Instead, the top quartile has a 54% chance of outperforming the average and the bottom quartile has only a 47% chance. That indicates greater gender diversity is correlated with a slightly greater chance of outperforming the average earnings margin.

However, McKinsey's chart magnifies the apparent effect of gender diversity with two framing devices. First, it doubles it by translating percentage points into percentages: the top quartile's 4 percentage point improvement over chance's 50% becomes phrased as an 8% improvement. Then McKinsey nearly doubles the effect again by comparing the top quartile to the bottom: 54 is 15% higher than 47. But comparison to the average is more relevant, since a company that does not actively prioritize diversity should end up around the average. The way McKinsey presents its data takes gender diversity's small improvement over chance and quadruples it.

In any case, the chart's first footnote admits that this 2014 data does not meet the standard threshold for statistical significance, a p-value below .05.

Notably, the studies never say what the earnings margin of diverse companies *is* relative to their industry average; it could be lower, for all readers know. It is not enough to know how *often* diverse companies outperform or underperform their industry average—we need to know how *much* they do so by. The scale of successes and failures matters, not just their probability. As London Business School finance professor Alex Edmans puts this critique, "The most standard way to do it is to say what's the average profitability of diverse versus the average profitability of non-diverse" companies.[6] Readers want to know how much money diverse and non-diverse companies make, but the studies occupy themselves solely with the question of how likely they are to make money. This approach is like comparing how likely lottery tickets are to win prizes without caring that some have a payout of ten dollars and others win a million.

Not only that, earnings margin only produces actual earnings when combined with revenue, a central aspect of business performance the studies never consider. The authors

[6] Edmans also rightly wonders why McKinsey carves its data into quartiles rather than some other division or none. Sarah Todd, "Is McKinsey Wrong about the Financial Benefits of Diversity?" *Quartz*, 29 July 2021, https://qz.com/work/2038103/is-mckinsey-wrong-about-the-financial-benefits-of-diversity.

surely had the data about executive diversity's relation to revenue, margin, and earnings growth, information that would strengthen their case if favorable, but chose to show one narrow measure of financial performance and hide the most fundamental ones.

But the biggest problem with McKinsey's argument is that it is entirely correlational. Boardroom diversity quotas and race and gender-based hiring and promotion policies are predicated on the belief that these kinds of diversity *cause* corporate outperformance. But the research they are based on explicitly disavows any causal claims.

McKinsey was upfront about this limitation in its summary of its first report. The second paragraph begins, "While correlation does not equal causation (greater gender and ethnic diversity in corporate leadership doesn't automatically translate into more profit), the correlation does indicate that when companies commit themselves to diverse leadership, they are more successful."[7] In other words, the correlation indicates a correlation. By the second report, the acknowledgment that correlation is not causation drops to the end.[8] In the third, this disclaimer no longer appears in the public summary at all. It is buried at the end of a 50-page version: "Correlation is not causation. There are real limitations, and we are not asserting a causal link. As with many levers of business performance, particularly at such a high level, this would be challenging to demonstrate, likely requiring detailed longitudinal studies. Yet, while not causal, the relationship is real."[9]

This is baffling. What does it mean for there to be a real but non-causal relationship between executive diversity and corporate performance? Without evidence that diversity causes outperformance, why should companies expect diversity measures to improve their performance? The relationship between ice cream consumption and murder is in a sense real, but entirely correlational; both happen more often during warm weather. Causation is what matters.

Even if all McKinsey's claims were true, this is hardly an analysis worth restructuring the corporate world over. The correlation between executive diversity and chance of earnings margin outperformance is minor, the relation to earnings itself is

---

[7] Dame Vivian Hunt et al, "Why Diversity Matters." *McKinsey & Company*, 1 Jan. 2015, www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/why-diversity-matters.

[8] Dame Vivian Hunt et al, "Delivering through Diversity." *McKinsey & Company*, 18 Jan. 2018, www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/delivering-through-diversity.

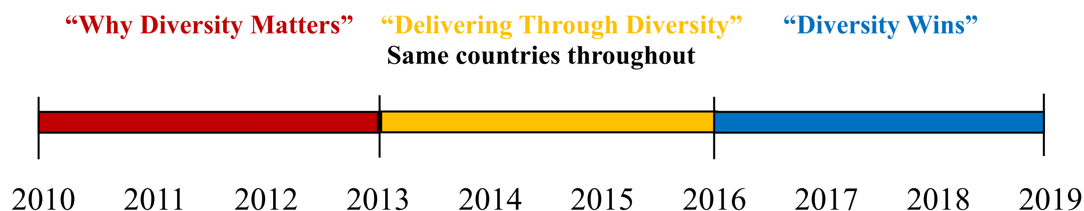
[9] "Diversity Wins," 51, https://www.mckinsey.com/~/media/mckinsey/featured%20insights/diversity%20and%20inclusion/diversity%20wins%20how%20inclusion%20matters/diversity-wins-how-inclusion-matters-vf.pdf.

unclear, and the authors explicitly refuse to assert any causal connection. But there are good reasons to think the correlation itself is just the product of a deeply flawed statistical methodology.

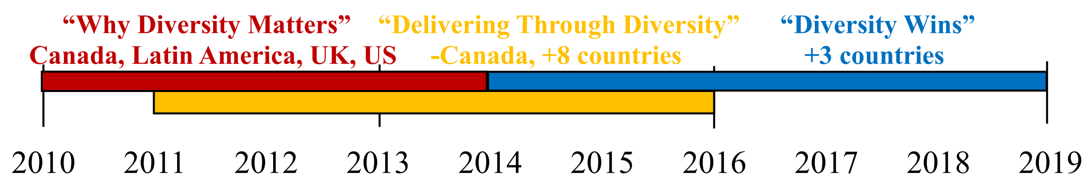
Cherry-picked data

McKinsey's hope is that the sheer robustness of its correlation across time and space can remove the need to prove causation. This is why its bar chart aims to demonstrate that the gap between the most and least gender diverse companies' chance of financial outperformance is widening over time. That argument requires a consistent underlying dataset—readers need reassurance that what is changing over time is the performance of the companies in the dataset, not the dataset's composition. Yet across the three reports, the underlying dataset is always shifting, displaying arbitrary choices in the timeframes it covers and companies it includes.

Here is what we might expect the reports' datasets on gender diversity to look like:



Instead, they look like this:



The choice of timeframes raises multiple red flags. The first report assesses EBIT margin over a four-year period, 2010-2013. The second and third each cover five years. Inexplicably, the reports do not cover *consecutive* time periods; instead, their datasets overlap in strange ways. The second report includes data from 2011 to 2015, omitting the first year of data from the prior report, including the other three, and adding two more years. Then the third report includes data from 2014 to 2018, omitting the first three years of data from the prior report, keeping the next two, and adding three more years of data.

This research methodology is highly suspect, especially when so many inconsistent decisions come with no justification and are disclosed in the fine print. It is also suspicious that after failing to find a statistically significant correlation in its first report's dataset, McKinsey adds more countries, removes Canada without comment, then finds one—it raises the question of whether Canada had less diverse companies that performed well financially. Cobbling datasets together in this patchwork way to show a strengthening correlation is arbitrary at best, dishonest at worst. It displays all the signs of p-hacking, the statistical trick of adding artificial criteria until one's data supports a predetermined conclusion.[10] Most people call it cherry-picking.

This multitude of basic methodological problems makes it all the more important to see whether independent researchers reach the same results as McKinsey; they do not. Texas A&M accounting professor Jeremiah Green and UNC business professor John Hand conducted a rigorous attempt to replicate McKinsey's analysis using S&P 500 data from 2015 to 2019. They found no statistically significant correlation between executive racial diversity and chance of EBIT margin outperformance. They also checked whether it had any correlation with EBIT margin itself, sales growth, gross margin, return on assets, return on equity, and total shareholder returns. They again found no statistically significant correlations.[11] McKinsey declined to comment on their results.[12]

The McKinsey studies do apply appropriate rigor at one stage: they group companies together by industry and geography and then measure how well diverse companies perform compared to the median EBIT margin of their peers.[13] But the fact that they correctly control for industry and geography when measuring *financial performance* makes it confounding that they choose not to when measuring *diversity*. They write, "For gender diversity, quartiles were based on the percentage of women at a given level, and set relative to the total ("global" sample) of 15 countries."[14]

That opens the door for successful industries and countries to skew the results. American big tech companies like Google and Amazon tend to have more executive gender diversity than Mexican textile companies. McKinsey's method of grouping them in the same pool means that its top quartile by diversity will be disproportionately composed

---

[10] This webcomic illustrates how one can use p-hacking to find a statistically significant correlation between green jellybean consumption and acne: https://xkcd.com/882/.

[11] Jeremiah Green and John R. M. Hand, "Diversity Matters/Delivers/Wins Revisited in S&P 500® Firms," 6 Aug. 2021, available at SSRN: https://ssrn.com/abstract=3849562 or http://dx.doi.org/10.2139/ssrn.3849562

[12] https://qz.com/work/2038103/is-mckinsey-wrong-about-the-financial-benefits-of-diversity

[13] "Diversity wins," 49.

[14] Id. at 48.

of successful Western companies. Only then does it control for industry and geography by measuring whether the Googles and Amazons outperform other American tech companies. Selectively choosing when to account for those important factors is a subtle but effective way of putting a thumb on the scale.

Once again, McKinsey measures something much stranger than one would expect. Readers want an answer to a simple question: do more diverse companies in a national industry make more money than less diverse ones? Instead, the studies investigate whether the most diverse companies in the world are likely to make more money than their national industry competitors—there is not even any proof they are more diverse than those competitors. This is an odd, Frankenstein-like combination of different questions one might care about, but it is not clear why anyone should care about this question itself.

Correlation and causation

Instead of diversity causing corporate outperformance, outperformance may cause diversity. McKinsey acknowledges this possibility at the end of "Diversity wins": "Just as we cannot assert causality, we cannot say definitively what drives the correlations we find. It is theoretically possible that the better financial outperformance enables companies to achieve greater levels of diversity. Companies that perform well financially may choose to deploy more of their resources toward more advanced talent strategies, thus allowing them to attract more diverse talent, for example."[15] Google and Amazon are good illustrations.

In fact, McKinsey's own methodology suggests corporate outperformance is more likely to be causing executive diversity. Professors Green and Hand repeatedly emphasize this problem: "Indeed, the structure of McKinsey's tests are such that by measuring firm financial performance over the four or five years leading up to the year in which they judge the race/ethnicity of firms' executives, the default direction of causality that McKinsey capture in the positive correlation they report is that better firm financial performance causes firms to diversify the racial/ethnic composition of their executives, not the reverse."[16]

McKinsey should have shown that corporations that had diverse executives went on to perform better, but instead it shows that corporations that performed better ended up having diverse executives. After its studies, as the economy sputtered and big tech companies laid off employees, DEI (diversity, equity, and inclusion) jobs have been hit hardest: one-third of DEI professionals have left their jobs within the last year, compared

[15] Id. at 51.
[16] Green and Hand, 16.

to a non-DEI attrition rate of 21%.[17] If successful companies thought these employees increased profitability, they would not be among the first to be fired. Some major companies conducting layoffs have also had sharp declines in racially diverse hires.[18] These facts suggest companies adopted diversity measures during boom times and are backing away from them now that profit is harder to come by.

Financial success may allow companies to spend money on unproven diversity practices that poorer companies cannot afford. The McKinsey study authors respond, "However, in practice this seems unlikely. We have observed that most companies only embark on a major transformation when they have a burning platform to do so."[19] But major companies do have a burning platform to embrace diversity measures—consultants have imposed it on them, along with bankers, politicians, lawyers, journalists, academics, and activists.

In addition to explaining why outperformance may cause diversity, this imperative creates a talent pool issue: as with university affirmative action, the top companies exhaust the pool of highly qualified minority applicants. That causes problems when all the other companies are forced to meet the same quotas. "Diversity wins" recognizes that diversity quotas may force STEM-related businesses to loosen their standards, given the limited pipeline of qualified candidates.[20] Lowering hiring standards could hinder operational performance. Performance could further suffer if employees lose motivation upon seeing peers receive career opportunities based on race or gender rather than merit.

This is why showing a mere correlation between diversity and corporate outperformance is not enough: people's careers are at stake. Jobs, promotions, and salaries are scarce resources. American capitalism promises workers that those finite goods will be distributed fairly—career opportunities are merited by those who would use them to maximize shareholder profit. That promise of fairness is broken when employers give out career opportunities based on tangential traits like race or gender. The subtitle of "Diversity Wins" gives a hint about which factors actually drive business performance:

---

[17] Kiara Alfonseca and Max Zahn, "How corporate America is slashing DEI workers amid backlash to diversity programs." *ABC News*, 7 July 2023, https://abcnews.go.com/US/corporate-america-slashing-dei-workers-amid-backlash-diversity/story?id=100477952#:~:text=The%20job%20losses%20owe%20to,DEI%20professionals%20told%20ABC%20News.

[18] Reyhan Ayas, Paulina Aceves, and Devan Rawlings, "Cutting Costs at the Expense of Diversity," *Revelio Labs*, 7 Feb. 2023, https://www.reveliolabs.com/news/social/cutting-costs-at-the-expense-of-diversity/.

[19] "Diversity Wins," 51.

[20] Id. at 42.

Diversity wins

How inclusion matters

As its title promises, though it spends most of its time talking about diversity, the report concludes that inclusion is what really matters.[21] "Diversity wins" elaborates, "Our evidence is that an emphasis on representation is not enough; employees need to feel and perceive equality and fairness of opportunity in their workplace. Companies that lead on diversity have taken bold steps to strengthen inclusion."[22] In other words, just as warm weather is the confounding variable that explains the correlation between ice cream consumption and murder, inclusion explains the weak correlation between diversity and outperformance. At the end of its trail of cherry-picking McKinsey arrives at the conclusion that companies do better when they treat their employees fairly and make them feel like they belong. On this, at least, Strive can agree.

But the big question is what diversity quotas have to do with inclusion; linking the concepts in acronyms does not connect them in reality. Treating employees equally and fairly is fundamentally at odds with distributing career opportunities based on race or gender. And it is hard to see how employees from disfavored groups can feel like they belong in a workplace if they see members of favored ones valued more highly there. This helps explain why the best-performing companies present in all three McKinsey reports had significantly *less* racial and gender executive diversity at the end than the beginning, a finding which puzzles it.[23]

We are left with common wisdom: businesses should treat their employees fairly and well. They should hire, promote, and reward those who contribute the most to shareholder profit. Race and gender appear to have nothing to do with it.

[21] "Diversity wins," 6.
[22] Id. in preface.
[23] Id. at 5.