GUGGENHEIM

June 2020

Hospital Emergency Room Wait Time Tracking During COVID-19 Pandemic

Shannon Callan, CFA
Data Consultant

Previously a fundamental research analyst, portfolio manager and director of equity research at the TCW Group, Shannon migrated to the field of data science. In recent years, she has worked with some of the world's largest asset managers to help them leverage alternative data in their investment research. Guggenheim Securities has contracted with Shannon's firm, Digital Trend Analytics, to provide data analytics support for its equity research team.

Key Message: Guggenheim Securities monitored the Emergency Room wait times of roughly 350 U.S. hospitals from late March to late May in an effort to evaluate the level of stress on the hospital system. Tracking the ER wait times hospitals posted on their websites provided an early read on the progression of the COVID-19 pandemic and confirmed trends seen in official data. Counterintuitively, ER wait times declined in locations where COVID-19 cases were increasing, signaling the extent of lost lucrative business for hospitals.

Introduction

In this time of data ubiquity, stories about applying data are highly relevant for analysts of all types. Stories about where data comes from, what the quality is like, and most importantly how it provides novel information are valuable because of the multitude of data odysseys analysts can embark on. Experiencing the pitfalls of early data exploration vicariously is generally preferred to doing it oneself, and lessons learned from one situation may be applicable to another.

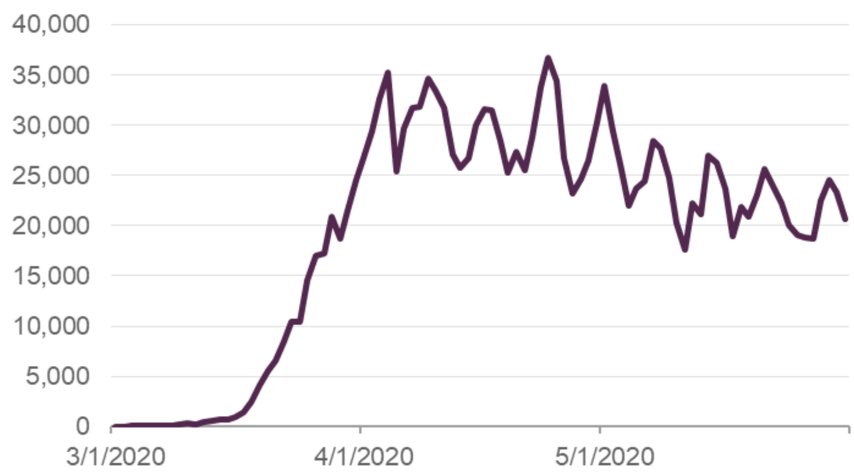
This is a story of a Guggenheim Securities data project that

- Was conceived and executed quickly in response to the COVID-19 pandemic.
- Leveraged a data source that was definitely "off the beaten path."
- Involved many unknowns and risks to its success.
- Yielded unexpected results.

Project Genesis

In mid-March, analysts and investors were scrambling to evaluate the economic fallout of the COVID-19 pandemic. Knowing that the amount of stress that the U.S. hospital system experienced would be a determining factor for the depth and length of the economic shutdown, Guggenheim Securities Senior Advisor on Healthcare Policy Neal Masia was seeking ways to track hospital conditions as the contagion spread. Masia noted that data on diagnosed cases, hospitalizations and fatality rates were not only backward-looking but were also subject to variability depending upon processes and definitions. In March, without any recent history of data collection during a pandemic in the U.S., it wasn't even clear that there would be reliable and timely statistics regarding diagnosed cases and fatalities.

New COVID-19 Cases Diagnosed in the U.S.



Source: Guggenheim Securities, LLC, Digital Trend Analytics, The New York Times

It occurred to us that a useful metric for monitoring the spread of COVID-19 might be the wait time at hospital emergency rooms. Many hospitals post their ER wait time on their website with most of those wait times based on internal hospital systems that track the time between arrivals and first encounters with a medical professional. By regularly consulting a set of these hospital websites, could it be possible to get a unique view of the level of stress on the U.S. hospital system?

Web Crawling, Web Scraping and Data Harvesting

A computer program that consults web pages on an automated basis and records relevant data points is commonly referred to as a web crawling or web scraping program. The project may be referred to as a data harvesting project. growing number of asset managers and investment banks seem to have found that data harvesting provides relevant data points such as product pricing and availability, local government statistics, and public commentary posted online. Project execution can be done in-house by IT professionals, or analysts skilled in computer scripting languages such as Python and R. Many IT services companies specialize in data harvesting as do many boutique data analytics service providers. Researchers can easily outsource all or part of a data harvesting project.

Project Risks

Every data project involves unique risks, which translate into costs, typically in the form of labor hours. A successful data project evaluates the potential gains from the valuable information the data will provide versus the cost of execution. Novel ideas such as the hospital ER wait time tracker are frequently riskier projects. In this case, the risks were weighted more towards the information payoff as the execution cost was low and adjustable depending on project scope. The information payoff was highly uncertain as there were open questions about the project's feasibility.

Were there enough hospitals to constitute a sample? Most hospitals don't post their emergency room wait times online and it was unclear if there would be enough of them that do to generate a representative sample.

How biased would the sample of hospitals be? While geographic biases could be mitigated, it is likely that there are quality differences between hospitals that post wait times online versus those that don't or lack the resources to do so. The hospitals that post their wait times on their websites are frequently ones that are more aggressively seeking new patients. There was a risk that in our quest to determine systemic stress we would be polling some of the least stressed members of the population. The question was how much the sample bias would distort the ground-truth picture.

Would ER wait times posted online reflect the demand for emergency room services due to the spread of COVID-19? It was possible that hospitals would stop posting their wait times or that they would stop updating them as the ER rooms became busier. Hospitals might add capacity to their patient intake processing as they became busier or process patients suspected of having COVID-19 in a way that wouldn't affect the ER wait time system's calculations.

What would be the characteristics of the data collected? Because there was no data set of hospital wait times readily available, key factors in data quality such as the volatility and amount of noise in the data were unknown. The optimal frequency of data collection was also unknown.

In addition to the unique risks mentioned above, managing web scraping programs is cumbersome. Websites frequently change in structure, and the process of accessing website information and recording it can fail for a variety of reasons. Missing data is a common problem in data harvesting projects.

Despite these risks, the surge in demand for data that could provide a read on the path of viral contagion was large enough that Guggenheim Securities decided to move forward with the project. At the time, investors knew that the economy was likely to drift into uncharted waters and that the right data could help them in these unusual circumstances. The surge in demand for data was matched with a surge in supply of data, but more data was considered better as each source could contribute to an information mosaic.

Because of the risks and unknowns, the key to the hospital ER wait time project's success was to gather just enough data that was of minimal quality in order to evaluate whether to continue building the program and improving the data collection process. There was a balance to keep between rapid execution, data quality and cost containment (hours spent on the project).

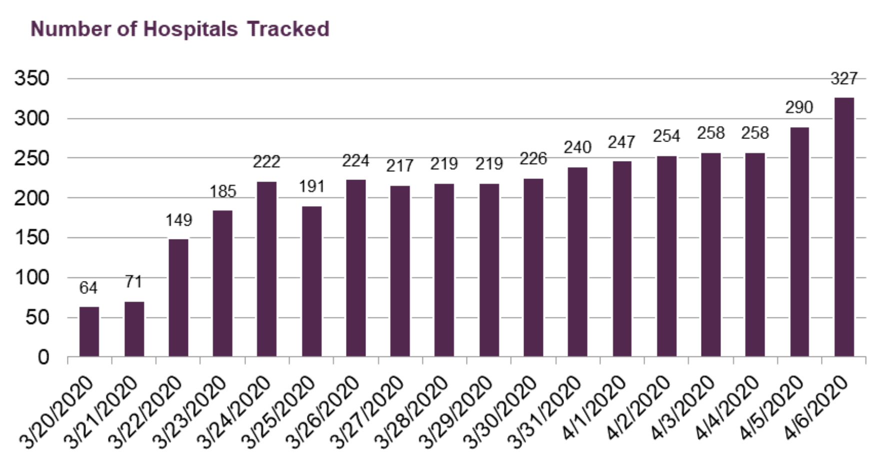
Project Execution

There are more than 6,000 hospitals in the U.S. Our analysis chose to focus on non-specialized hospitals, excluding pediatric/children's hospitals, surgical specialty hospitals, and ones for specific illnesses such as cancer or injuries such as burns. We attempted to select a group of hospitals that proportionately represented regional population distributions.

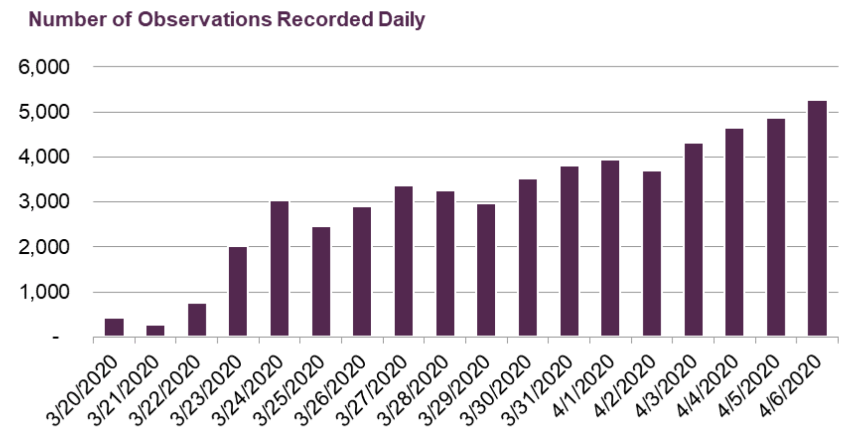
We began to collect data as soon as hospitals and URLs were identified. More data was collected than used due to data quality issues, such as hospitals that published "NA" too frequently. By mid-April, sample size stabilized at roughly 325 hospitals, representing more than 5%, but less than 10%, of the relevant hospital population (non-specialized hospitals with emergency rooms). Was it enough? Traditional statistics bases the optimal sample size on the margin of error and sampling confidence level. With those being unknown, a 10% sample size is often a data practitioner's rule of thumb for optimal size. While the project was still in a proof of concept stage, additional time spent on growing the sample size was not a justifiable endeavor, particularly as data cleaning and analysis needs grew.

The Web Crawling Programs

The programs that collected ER wait times from the hospital websites accessed just under 80 different URLs as some public websites listed multiple hospitals. Roughly half of the websites utilized JavaScript in their wait time display while the other half used basic HTML/CSS structure. This made it technically relatively easy to collect and record data in spreadsheets.



Source: Guggenheim Securities, LLC, Digital Trend Analytics



Source: Guggenheim Securities, LLC, Digital Trend Analytics

The data collected never became “big data.” Hourly data collection was targeted from roughly 7 a.m. to just after midnight ET. Running 18 collections daily for roughly 350 hospitals added 6,300 rows per day to the dataset. Delivery of the entire dataset could be done in a Microsoft Excel spreadsheet. While the data could be analyzed in Excel, analytical programs such as R, Python and Tableau were more effective.

Each hospital in the dataset was associated with an owner/operator, zip code, metropolitan area, and a unique identifier. The primary KPIs listed below could then be viewed at different levels of aggregation.

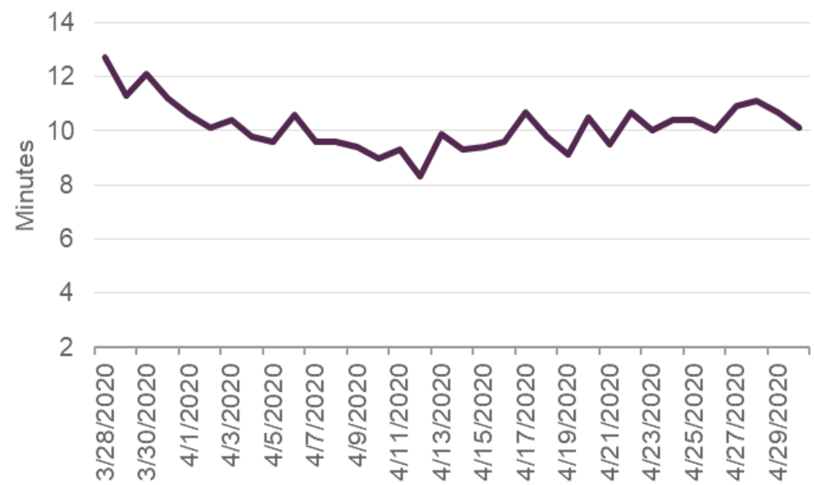
- Average wait time.
- % of hospitals with wait time less than 15 minutes.
- % of hospitals with wait time longer than 30 minutes.

Ideally these KPIs would have been displayed in an interactive dashboard that is updated in real time. Instead, the data was analyzed in batches and summarized in Excel spreadsheets. The Excel spreadsheets were needed in the early stages of gathering and reviewing data, and a crucial need to migrate to a dashboard never appeared. While Guggenheim Securities’ clients had the option of working directly with the data themselves, the data was mostly used to support Guggenheim Securities’ Healthcare research.

Hospital ER Wait Time Tracker Results

As data collection began in late March, it became evident that hospital ER wait times were decreasing. This was the opposite of what we had expected, and the trend continued into mid-April.

Average Hospital Emergency Room Wait Time



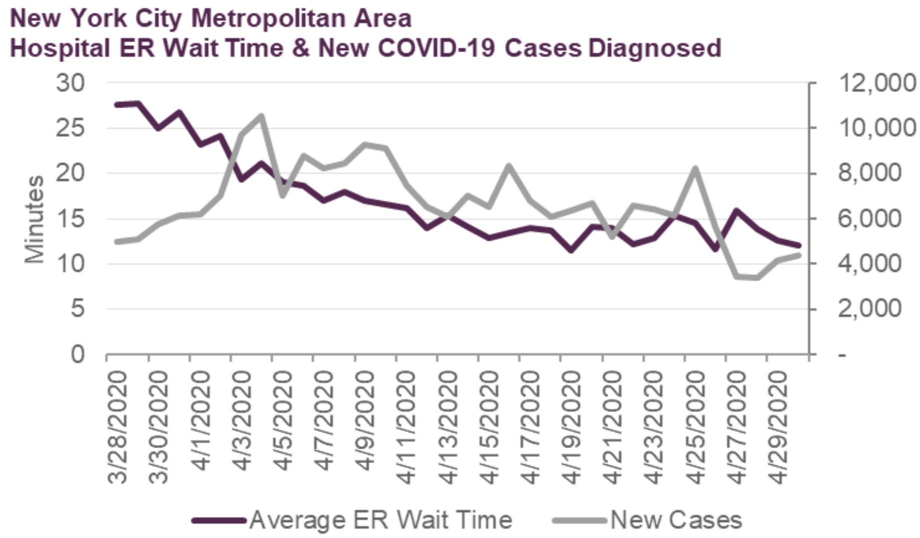
Source: Guggenheim Securities, LLC, Digital Trend Analytics

Shorter wait times were a result of fewer ER visits. For those not suffering from COVID-19, major contributing factors to reduced ER visits were:

1. Fewer accidents occurring as people stayed home and were less active. According to the CDC, the US has roughly 139 million ER visits per year, accounting for over 16 million annual hospital admissions out of 36.5 million total admissions. Over 40 million ER visits are injury-related. This implies daily ER visits of roughly 381,000, with roughly 110,000 due to injury. The general slowdown in daily life presumably reduced the daily toll significantly, creating capacity for COVID-19 patients. In early April there were roughly 16,000-17,000 patients hospitalized in the New York City area for COVID-19, according to The COVID Tracking Project.
2. Avoidance of hospital ERs due to fear of virus contagion. If medical care was needed, people made efforts to get that care in ways they believed were less likely to result in contagion.

For lower income people suffering from COVID-19 symptoms, a variety of impediments may have kept them from seeking help from a hospital ER. Deaths at home spiked in the New York City area and in other areas with high levels of contagion, with the increase attributable to individuals suffering from COVID-19 and those needing medical care but staying away from hospitals (https://www.nytimes.com/interactive/2020/04/21/world/coronavirus-missing-deaths.html).

In early April, the rapid decline in hospital ER wait times was accompanied by stories in the media of New York City metropolitan area hospitals and healthcare first responders being overwhelmed by COVID-19 cases. In retrospect, we can see that new cases in the area were plateauing in the first half of April and then began declining.

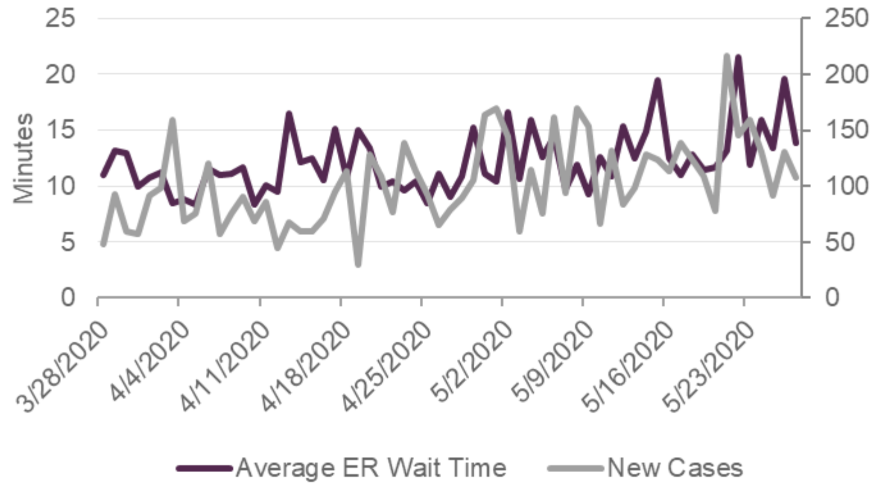


Source: Guggenheim Securities, LLC, Digital Trend Analytics

The data harvesting project achieved its goal of providing insight into the financial impact of COVID-19 on the hospital system. Two crucial areas of revenue for hospitals are a) elective surgeries and b) emergency room visits that lead to inpatient admissions. Loss of revenue from ER admissions generated an acute financial burden for most hospitals, even in areas with fewer diagnosed COVID-19 cases. If healthcare analysts were prone to underestimating the decline in high margin hospital revenues, and sales of companies that supply to this segment, this data would have provided reason to pause and reconsider.

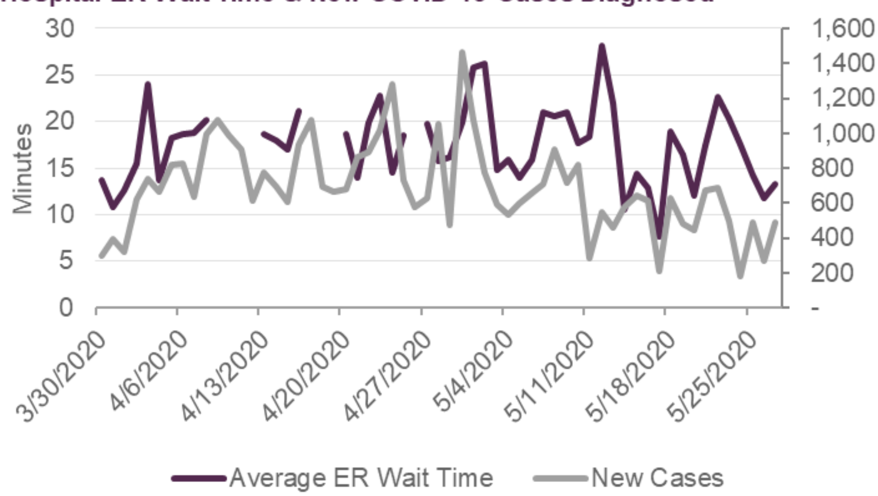
However, the goal of gaining insights into the spread of COVID-19 was only partially achieved with a mixture of hits and misses. The two opposing factors, COVID-19 illnesses on one hand and reduced accidents combined with hospital avoidance on the other, affected ER wait times differently and they shifted in importance over time. The data seemed to fit with publicly released data on new cases diagnosed, but rather loosely. Ohio cities Cleveland and Cincinnati are good examples, as is Philadelphia, shown in the charts below.

Cinncinatti and Cleveland, Ohio
Hospital ER Wait Time & New COVID-19 Cases Diagnosed



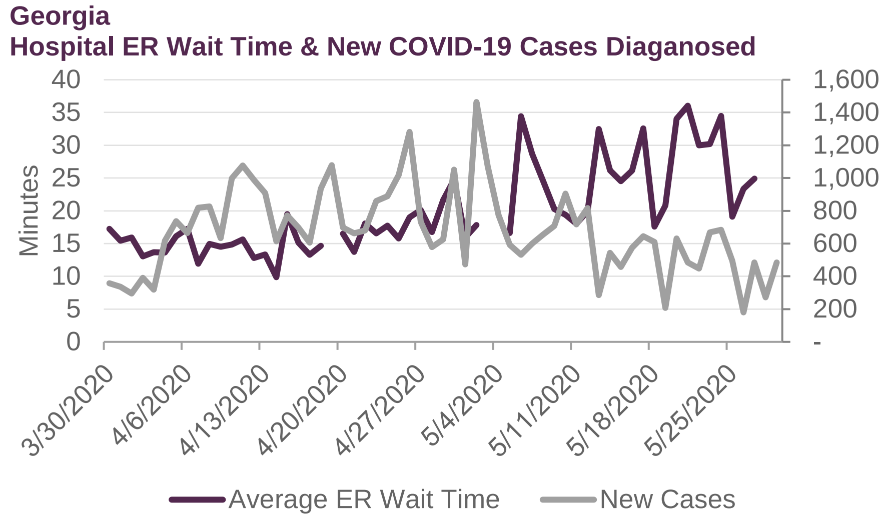
Source: Guggenheim Securities, LLC, Digital Trend Analytics

Philadelphia Metropolitan Area
Hospital ER Wait Time & New COVID-19 Cases Diagnosed



Source: Guggenheim Securities, LLC, Digital Trend Analytics

In Philadelphia, new cases peaked in late April and wait times then declined in tandem with new cases. Meanwhile in Georgia, new cases peaked on April 30th and average ER wait times subsequently increased.



Source: Guggenheim Securities, Digital Trend Analytics.

In South Carolina, a state that has had one of the shortest stay-at-home orders, average ER wait times were stable at around 20 minutes in early April but then gradually rose to averaging 30-35 minutes in the last half of May.

Data Value

The hospital ER wait time dataset was valuable in that:

1. It pointed out stresses in the hospital system that were unexpected, providing investors in the healthcare sector with a better understanding of the financial difficulties hospitals are facing.
2. It provided an early read from the front lines of the COVID-19 pandemic at a time when there was a high degree of uncertainty about the future rate of contagion and how it might spread geographically. Ultimately, numerous reputable data sources such as WHO, CDC, The COVID Tracking Project, and Johns Hopkins CSSE emerged to provide timely data on COVID-19 cases. Numerous efforts, many of them non-profit, built programs to aggregate data from both larger and smaller, often local, data sources.
3. It confirmed the official government data with an unrelated set of observations. While there is currently a high degree of confidence in the data on diagnosed cases and fatalities, that was not the case at the beginning of the crisis. An unrelated data source that moves in line with a bellwether data source, as is the case with the hospital

ER wait time data, provides a cushion of comfort for using the bellwether data source.

As time passed, however, the utility of the information the data provided declined and the project was discontinued at the end of May. The social unrest in June and subsequent opening of the economy has since shifted the spotlight back onto hospital resource constraints. Depending on the level of interest in hospital ER wait time tracking, the data harvesting program may be restarted.

Disclaimer

This Commentary is intended for Institutional investors only and not a product of GS Research Dept.

This institutional communication has been prepared for your informational purposes by Guggenheim Securities LLC (Guggenheim) and its Consultant partners, and is not a product of Guggenheim Securities' Research Department. This material should not be construed as a research report, recommendation or advice, or an offer or solicitation with respect to the purchase or sale of any investment. The information in this communication is not intended to provide a basis on which you could make an investment decision on a traded security or its issuer. The reader should conduct his/her own independent analysis before making any investment decision. Any discussion of prices or yields are historical or indicative and should not be interpreted as bids or offers. Past performance is not indicative of future performance. Any forward-looking statements are valid only as of the date they are made.

The information herein is believed to be derived from reliable sources, but the information is not guaranteed as to accuracy and does not purport to be complete. The information may be subject to change without notice, and Guggenheim makes no representation and gives no warranty that it is accurate or complete. Guggenheim undertakes no obligation to provide you with any additional or supplemental information or any update to or correction of the information contained herein. Guggenheim, its officers, employees, affiliates and partners shall not be liable to any person in any way whatsoever for any losses, costs or claims for your reliance on this material. Guggenheim may have a business relationship with an issuer mentioned, or may otherwise underwrite, hold a position, buy or sell as principal in the securities of any issuer mentioned. With respect to any position Guggenheim may hold, Guggenheim may have acquired any such position at prices no longer available and may have interests different from or adverse to your interests.

This material is distributed with the understanding that it does not provide accounting, legal or tax advice; you should consult appropriate advisors concerning such matters. This material is not to be reproduced or redistributed to any other person outside of your organization or published in whole or in part for any purpose absent the written consent of Guggenheim. The sender does not accept liability for any errors or omissions in the content which arise as a result of e-mail transmission.