Getting from big data to analytics

Four challenges for health care entities

EY
Building a better working world

Are you sick of hearing about big data yet? The term went from obscure to ubiquitous so rapidly that it often engenders a good deal of eye rolling in health care circles. Yet, behind the hype and the armies of consultants bearing PowerPoint decks lies a powerful — and largely valid — premise.

The basic idea can be summed up as follows. Health care systems — creaking under the weight of unsustainable cost inflation — desperately need cost-efficient solutions. These budgetary pressures are already fueling a shift from fee-for-service to pay-for-performance across many markets. At the same time, we are witnessing an unprecedented explosion in the volume, variety and velocity of health care data: everything from electronic health records to payer claims data and even the real-time information generated by mobile health technologies. What if we could combine these data streams to get the "big picture"? More robust data, for instance, could be used to identify and reward the most effective interventions — the critical foundation underpinning the move to pay-for-performance. Similarly, real-world, real-time data could be used to understand how behavioral and environmental factors influence health outcomes — vital for stemming the rising tide of chronic disease costs. Or consider the ultimate magic bullet: the promise of predictive and prescriptive analytics, which involve using data analytics to identify the small percentage of individuals most likely to drive costs and intervening before their high-cost events occur.

This is a compelling vision and one that is sorely needed to address the biggest challenges of health care. But, as in most things, details matter. Even as big data creates new ways of tackling health care challenges, it also raises new hurdles. Payers and providers looking to harness the potential latent in this information will need to address these issues — which we have grouped into four key challenges.

Challenge **one.** Getting along well with others: combining data

As articulated above, the potential of big data lies in the ability to connect dots across different data streams. But achieving this vision runs headfirst into the first big challenge: gaining access to other data streams. Consolidating data has never been easy because of the fragmented and localized nature of health care delivery, but the challenge has been compounded in a couple of ways by the arrival of big data.

First, big data raises **technological hurdles**. The explosion of data streams means entities have to deal with much larger volumes of data, often in diverse and dissimilar formats and including structured and unstructured data. The data involved can extend well beyond health care data as traditionally defined (health records, payer claims, pharmacy data, etc.) to include things like lifestyle or financial data (spending patterns, credit history, etc.). And, while we haven't yet seen much consolidation of mobile health data from wearables, apps and devices, we expect this to change soon as these sources of information become more widespread and reliable.

The technological challenges in combining these data sources are very addressable — but it is critical for data to be structured in ways that allow for integration and robust analysis. As information from multiple sources is combined, for instance, analysts need to pay careful attention to indexing the data and setting up the structural relationships between different fields that provide context and allow for richer analysis.

In addition to these technological concerns, the arrival of big data makes accessing data more challenging because of **strategic considerations**. As organizations realize that data is now one of their most valuable assets, they are becoming even more circumspect about when and how they share it. Indeed, the business models of many new digital health start-ups are based on monetizing data.

This challenge is compounded by concerns about **privacy** regulations (e.g., the Health Insurance Portability and Accountability Act, or HIPAA, in the US). Organizations often over-comply with privacy laws, assuming them to be more restrictive than they actually are. The net effect is that entities become overly conservative about sharing data.

The strategic and privacy concerns can be addressed through careful negotiation as organizations define the terms under which information is shared and how value is allocated between participants. Since this is often a contentious process, it can help to have a neutral third party handle the most controversial aspects. For instance, EY recently helped set up a Beacon Community in a US state to improve Type 2 diabetes outcomes in the region through the use of electronic health records. Several hospital systems and insurance companies had to share data to create a meaningful picture of the population's health. This was a challenging process and the participants ultimately reached agreement by having a third party house the data. Each organization's data was de-identified and added to the central repository. This allowed the consortium to analyze the data and address diabetes from a population health perspective. In addition, this arrangement also allowed participants to make targeted interventions at the individual patient level — more on this in the next section.

Challenge **two.** Garbage in, garbage out: data quality

Addressing the first challenge and gaining access to different data streams allows an entity to get from fragmented data to big data. Before the data can be useable, though, organizations also need to ensure that it is of high quality — the second big challenge. The issue of data quality — much like the challenge of accessing information — is a perennial problem that has always been part of statistical analysis, but it is compounded by the arrival of big data.

For instance, the huge increase in the volume of information from multiple organizations increases the likelihood of **data inconsistencies,** which could skew analyses and produce spurious findings. Interestingly, while big data has compounded this challenge, it has also produced potential solutions. For instance, the ability to access large volumes of information from multiple sources now enables organizations to use master data management to remove inconsistencies.

Consider the example of a hypothetical patient named Andrea J. Smith, who refers to herself as "Ann" in most day-to-day interactions. Imagine this patient has records in three different health organizations. Her name was entered as "Andrea Smith" in the first organization and as "Andrea J. Smith" in the second. In the third, her name was entered as "Ann Smith" and "Anne Smith" on two different occasions. Master data management allows practitioners to look for commonalities across large numbers of data fields to infer that these different records refer to the same individual, and thereby create one master record for the individual. With the arrival of big data, analysts can look at hundreds of data fields to make determinations that are very accurate.

A second complication with health care data is that entities are often dealing with **de-identified information**. In many markets, privacy regulations require health care entities to remove patient identifier information before data is shared. Without good workarounds, this would significantly limit the use of data, by making it harder to combine data from different sources. Luckily, there are workarounds for this challenge. Techniques such as master data management allow data to be combined even when it is de-identified — analysts can use the abundance of other variables to match up de-identified data from different sources. This allows them to know that the data refers to the same person, even though they may not know the individual's name.

De-identified data could also make it harder to conduct predictive and prescriptive analytics. Analytical methods might successfully identify a patient at risk of an adverse event, but this information is only useful if providers can get to the patient in time and intervene in a preventive way. With de-identified data, however, practitioners may have no way of getting to such patients since they cannot identify them.

Again, there are workarounds to this challenge. First, in cases where practitioners cannot target specific individuals by name, they can still focus interventions on targeted subpopulations of at-risk patients — for instance, by identifying postal codes or metropolitan statistical areas where large percentages of individuals are at risk. Second, contractual arrangements can be constructed to allow data to be shared in de-identified ways, while still enabling targeted interventions at the individual level. For instance, in the example cited earlier of the diabetes Beacon Community, the neutral party housing the data obtained de-identified data from participating hospitals and insurers. While the consortium could only conduct analytics at the population level, the third party was also able to provide information about specific high-risk individuals back to the original holder of that person's record. That organization could then identify these individuals by name and intervene with them in a timely way.

Challenge **three.** Data needs adult supervision: analytics and algorithms

The early hype around big data often envisaged a future in which large amounts of data would simply crunch itself. A 2008 cover story in *Wired* magazine titled "The End of Theory" posited that, with vast amounts of data, we would no longer need the long-standing scientific method — generating and testing hypotheses. Instead of hypothesis testing, *Wired* argued, "petabytes allow us to say: 'Correlation is enough.'"

The poster child for this worldview was Google and — in an implementation that hit very close to home for health care — the Mountain View-based company's Google Flu Trends website. The premise behind Google's approach was as compelling as it was intuitive. The company figured out that when people start feeling sick, one of the first things they do is go to the web and search for information on their symptoms. By looking for spikes in web searches for flu symptoms, Google was able to accurately predict influenza outbreaks and, even better, was able to do so weeks ahead of the scientists at the Centers for Disease Control and Prevention.

Google Flu Trends embodied everything that this vision of big data represents. It was built not by professional epidemiologists, but rather by a small team of computer programmers. Their methodology didn't need any scientific background or context. It didn't need to build any models. It didn't need to generate hypotheses and test them. It merely looked for telltale patterns in web searches and — armed with that information — generated accurate, real-time forecasts.

Until it didn't. In 2013, Google Flu Trends stumbled. Its forecasts became considerably less accurate. What went wrong? One explanation is that the 2012-13 flu season had been particularly severe, which had generated a lot of media coverage of the flu. This, in turn, led to a larger-than-expected increase in web searches as more people started looking for information even though they weren't exhibiting symptoms. Google's algorithm failed to account for this behavioral shift, and overestimated flu cases in its forecasts.

This is just one anecdote, but it illustrates a larger point. The simple, correlation-based, let-the-data-run-itself vision of big data turns out to be just that — simplistic. Big data *doesn't* run itself. To the contrary, it needs adult supervision. Indeed, dealing with large amounts of data may require even *more* care and human intervention than statistical analysis of smaller data sets. With vast volumes of data and large numbers of independent variables, the likelihood of finding spurious correlations increases sharply. Outliers can skew results, and algorithms, left unattended, can run wild.

The solution, as you might expect, is to bring together the best of machine learning and human oversight. Today's analytical techniques start with algorithms designed by data scientists with large amounts of medical knowledge and context. Algorithms have tremendous knowledge built into them about symptoms, disease progression, comorbidities, drug interactions and more. Consider diabetes, one of our biggest public health challenges and a very complex disease.

Diabetes patients can show up with any number of early symptoms ranging from blood clots to heart problems and circulatory issues. Over time, they may develop comorbidities, which can evolve in different ways and have to be managed carefully. The people building algorithms — not to mention the algorithms themselves — need to understand and account for this sort of context.

Even more important, these algorithms aren't static. Through the use of artificial intelligence and machine learning, they adjust over time, in response to new data and changing conditions. Here, too, adult supervision is required. Left to itself, an algorithm may fixate on a red herring and progress down an irrelevant path. Human analysts can run sanity checks on algorithms as they evolve and can nip spurious findings in the bud.

Challenge **four.** Moving at the pace of Moore's Law: focus and flexibility

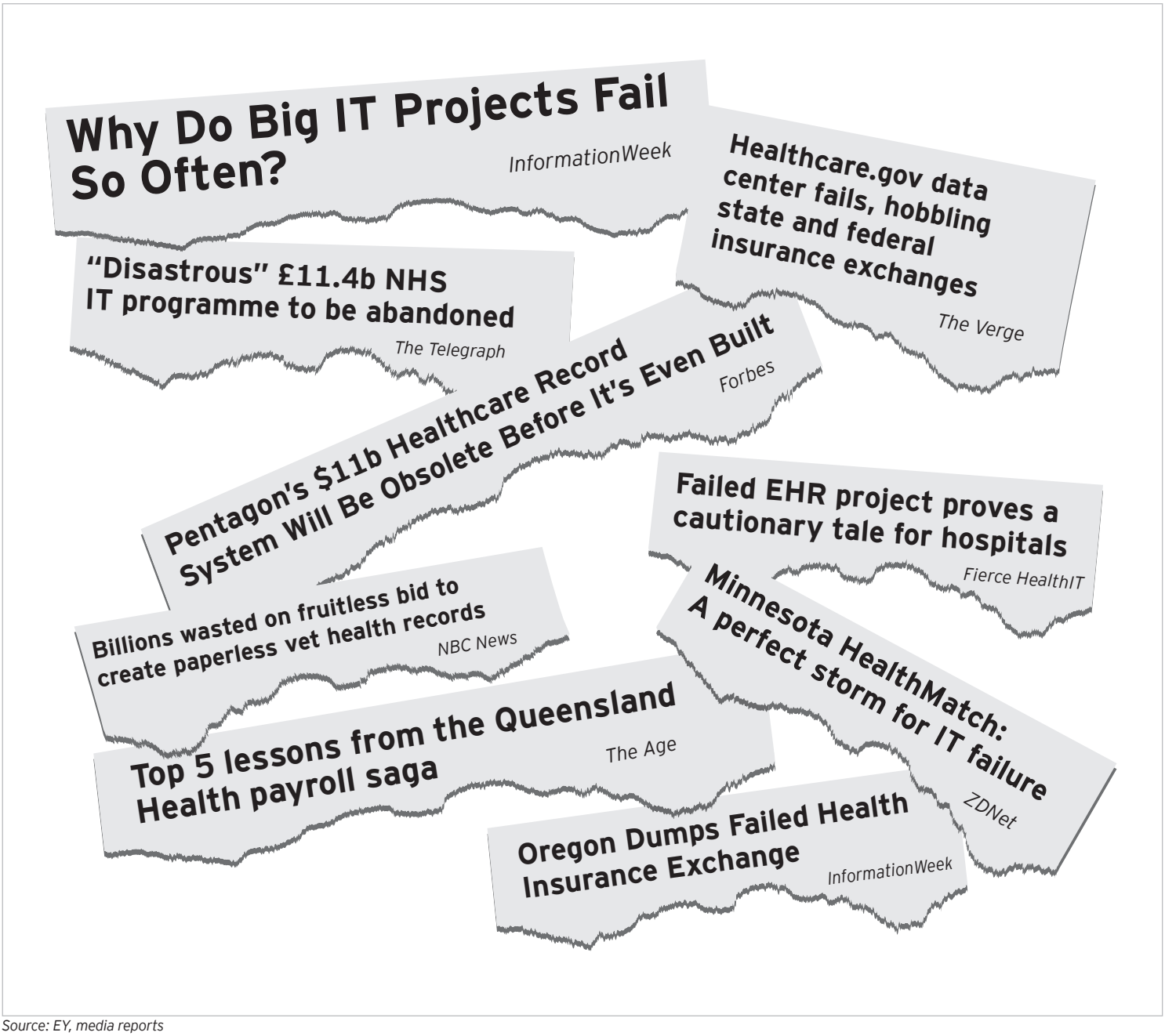
Beyond the challenges discussed previously — accessing and integrating data, cleaning and structuring it, constructing robust algorithms and analytics — organizations face a fourth challenge that is more fundamental and pervasive. After all, health care analytics aren't something that is done one time in a static and unchanging world. To the contrary, entities are operating in a space that is constantly in flux — as organizations enter and exit partnerships, work with new sources of data, and respond to changes in information technology (IT) standards and capabilities. The question for health care organizations, therefore, is: how do you invest in health care analytics in ways that are flexible and scalable?

To appreciate the extent of the challenge, consider how much the space of data analytics has changed in just the last five years. Five years ago, the term "big data" barely existed. In the US, the HITECH Act had recently been passed and provider organizations were just starting to think through their electronic health record implementation. Since then, we have seen dramatic increases in the volume and variety of health care data and vast improvements in the technologies and analytical techniques available to make sense of this information.

Consider also that health IT investments tend to be expensive projects that can take months, if not years, to implement. Given the rapid clip at which technology standards and platforms change — and the fact that health care entities are adjusting to major reforms in many markets — organizations run the all-too-real risk that by the time a project is completed, it may have been overtaken by events on the ground and may no longer meet their needs.

Lastly, consider that analytics is not a core competency for most health care entities. While the world of data analytics moves at the speed of Moore's Law, health care has historically evolved at a glacial pace. Furthermore, data analytics requires complex skills, specialized knowledge and deep experience that many in health care lack. The all-too-common result is health IT projects that run late, go over budget, and fail to deliver promised results and capabilities. Occasionally, this happens in a spectacularly public fashion, such as the fumbled launch of the Healthcare.gov website in the US and the

Exhibit 1. No news is good news? Health care IT investments often go awry



Source: EY, media reports

NHS' scrapping of a £12 billion electronic medical record project in the UK — considered the most expensive IT project failure of all time. More often, it happens without the attention-grabbing headlines, but with painful consequences for health care organizations that are saddled with large sunk investments in non-functional assets.

So, how do you proceed? A strategic approach to health analytics involves two elements:

- **Focus:** since analytics requires complex skills and large investments, a smart strategy is to focus only on functions that are true core competencies.
- **Flexibility:** a rapidly changing technological and business environment requires responses that are flexible and scalable.

As you formulate your analytics strategy, evaluate your options with these two elements — focus and flexibility — in mind.

For instance, one option is to "go it alone." A number of health care entities choose to make large investments in building infrastructure and capabilities in-house — either organically or through acquisitions. While some organizations may find success using this tactic, it involves large up-front investments and considerable risk that, even after spending hundreds of millions of dollars, they may end up with data warehouses that are unusable. Such approaches are unfocused (taking health care organizations well beyond their core strengths) and inflexible (building expensive infrastructure internally and limiting companies' ability to change tack in response to evolving market conditions).

More strategic responses give organizations greater focus and flexibility. This often involves working with third parties, for instance through strategic alliances in which others conduct data analytics. In addition, organizations are increasingly approaching data analytics as a service — borrowing a page from the business process outsourcing trend that has become commonplace in many business functions. Indeed, we would argue that this approach makes even more sense with respect to data analytics than with respect to functions such as human resources or finance, since analytics is further removed from health care entities' core strengths.

The "as-a-service" approach can be applied to various aspects of data analytics. Some organizations may adopt software-as-a-service or platform-as-a-service models, in which they rely on could-based programs or platforms but still conduct analytics internally. Such approaches give entities more flexibility, since they can scale their use of these IT assets up or down based on changing needs. However, these models are still relatively unfocused, since organizations have to build internal capabilities to conduct analytics, which is outside their core strengths.

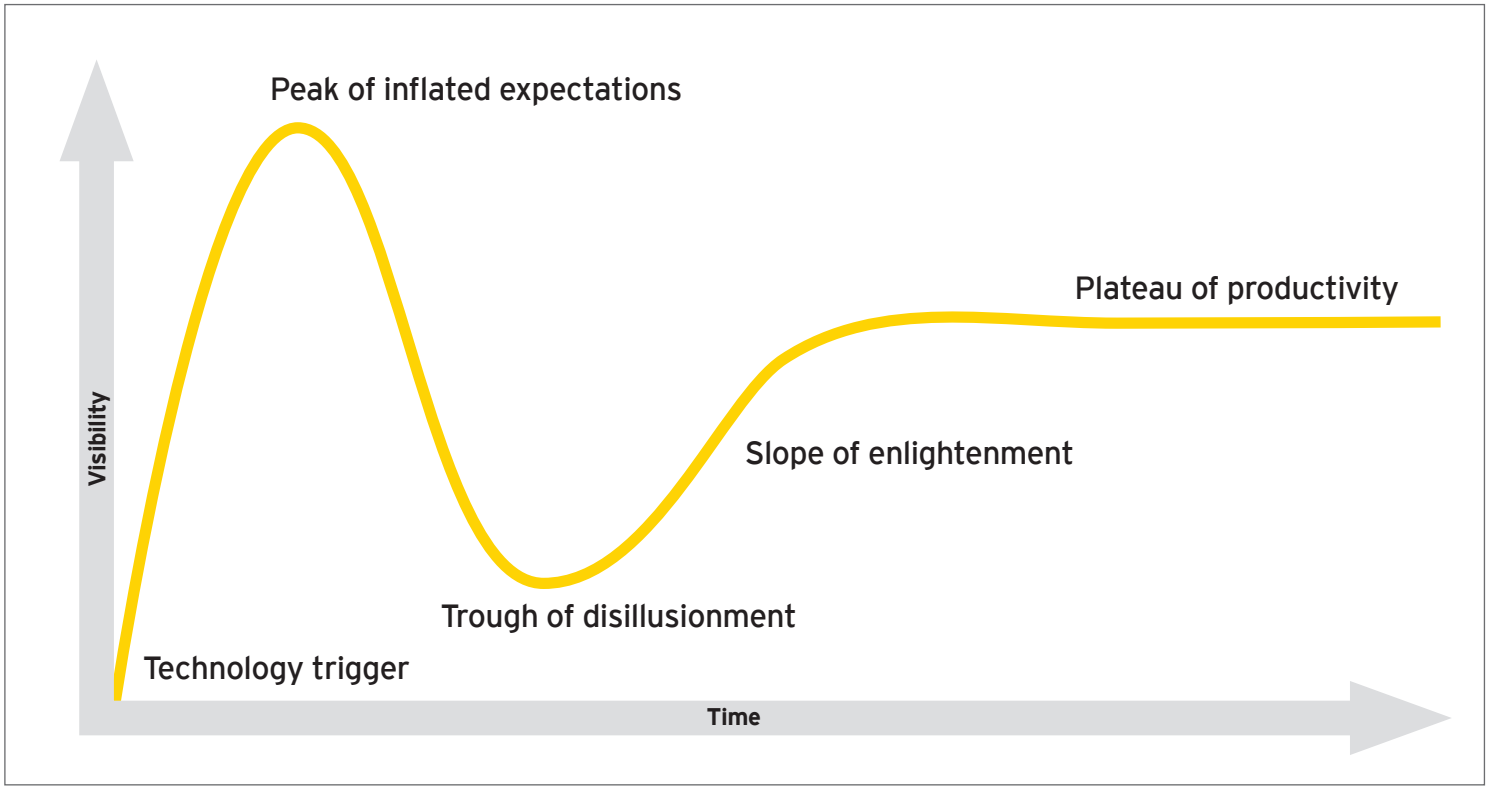
The most focused and flexible approach is an analytics-as-a-service model, in which organizations rely on third parties for all their analytics needs. Instead of spending hundreds of millions up front on relatively risky investments with uncertain payback, subscribers to a service pay only a monthly subscription or an annual fee. Through contractual arrangements, the accountability for providing analytical insights (and, therefore, risk) is shifted to service providers. And instead of waiting years while internal infrastructure and capabilities are built, companies can get to actionable insights very quickly.

Moving ahead

As you think about the potential of big data for your own organization, it may be worth remembering the Hype Cycle developed by Gartner, a technology research firm. The Hype Cycle includes five phases that new technologies go through. It's easy to see how this pattern is playing out in data analytics. Interest in big data has already been through the "peak of inflated expectations" and is now languishing at the "trough of disillusionment." It's worth keeping in mind, however, that neither the peak nor the trough accurately represents the long-term potential of big data. The truth lies somewhere in between.

To get to Gartner's "plateau of productivity," organizations need to unleash the potential of big data through meaningful analytics. We've seen individual health care entities fall into their own troughs of disillusionment by making large internal investments in health IT that extended organizations far beyond their capabilities and failed to pay off. It's critical, therefore, that you look for approaches that allow you to focus on what you do best and give you the flexibility you need in a rapidly changing market.

Exhibit 2. The Gartner Hype Cycle



Source: Gartner

Contacts

Jacques Mulder
Global Health Sector Leader
+1 212 773 1499
jacques.mulder@ey.com

Bill Fera
Global Health Analytics Leader
+1 412 644 0551
bill.fera@ey.com

Jim Costanzo
Global Health Advisory Services Leader
+1 248 705 0441
jim.costanzo@ey.com

Gautam Jaggi
Health Reimagined Thought Leadership Director
+1 617 374 6315
gautam.jaggi@ey.com

EY | Assurance | Tax | Transactions | Advisory

About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

About EY's Health Reimagined initiative
Health care is in the midst of a once-in-a-lifetime transformation that is blurring traditional boundaries and redefining health care as we know it. The move to "Health 2.0" — the patient-centric, technology-enabled, prevention-focused future of health care — is being catalyzed by nontraditional entrants from far-flung sectors, including technology, telecommunications, retail trade and more. These disruptive forces demand bold new approaches, alliances and business models.

EY's Health Reimagined initiative is a cross-sector program that brings together professionals and perspectives from multiple industries to develop insights and solutions that are aligned with the future of health. We can help you navigate your way forward and achieve success as you transition to Health 2.0.

EYG no. GC0011
CSG No. 1510-1733832

ED None

ey.com/healthreimagined