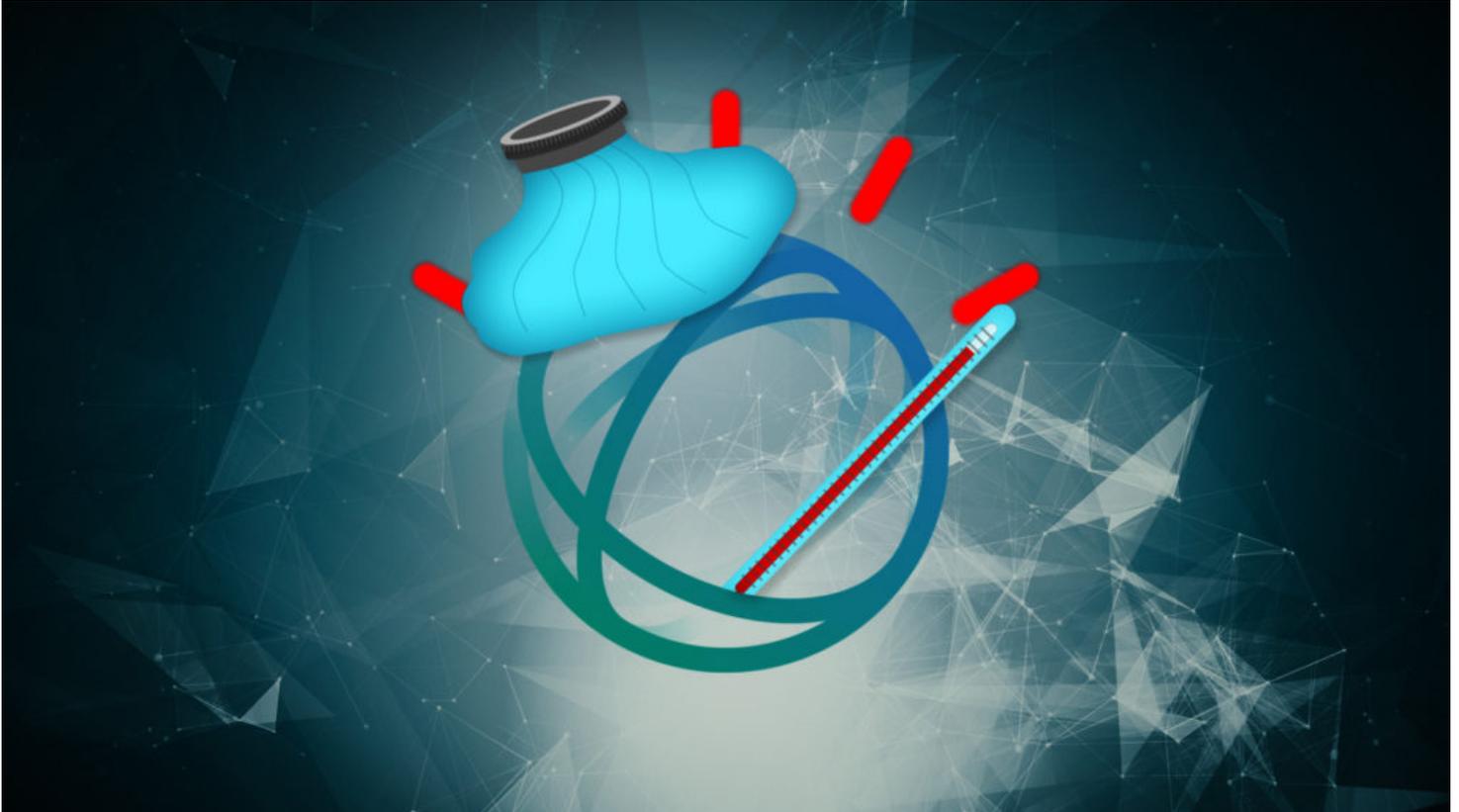


## IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show

By [Casey Ross @caseymross](#) and [Ike Swetlitz @ikeswetlitz](#)

July 25, 2018



Alex Hogan/STAT

Internal IBM documents show that its Watson supercomputer often spit out erroneous cancer treatment advice and that company medical specialists and customers identified “multiple examples of unsafe and incorrect treatment recommendations” as IBM was promoting the product to hospitals and physicians around the world.

The documents — slide decks presented last summer by IBM Watson Health’s deputy chief health officer — largely blame the problems on the training of Watson by IBM engineers and doctors at the renowned Memorial Sloan Kettering Cancer Center. The software was drilled with a small number of “synthetic” cancer cases, or hypothetical patients, rather than real patient data. Recommendations were based on the expertise of a few specialists for each cancer type, the documents say, instead of “guidelines or evidence.”

STAT has seen portions of the two presentations, from June and July of 2017. At the time, they were shared widely with the management of IBM’s Watson Health division. The documents contain scathing assessments of the Watson for Oncology product by customers and conclude that the “often inaccurate”

recommendations raise “serious questions about the process for building content and the underlying technology.”

IBM has not publicly acknowledged the shortcomings of the software, which uses artificial intelligence algorithms to recommend treatments for individual patients. To the contrary, top company executives told customers and others that Watson for Oncology’s advice for physicians is based on data from real patients, and that it had won nearly universal praise from doctors around the world.

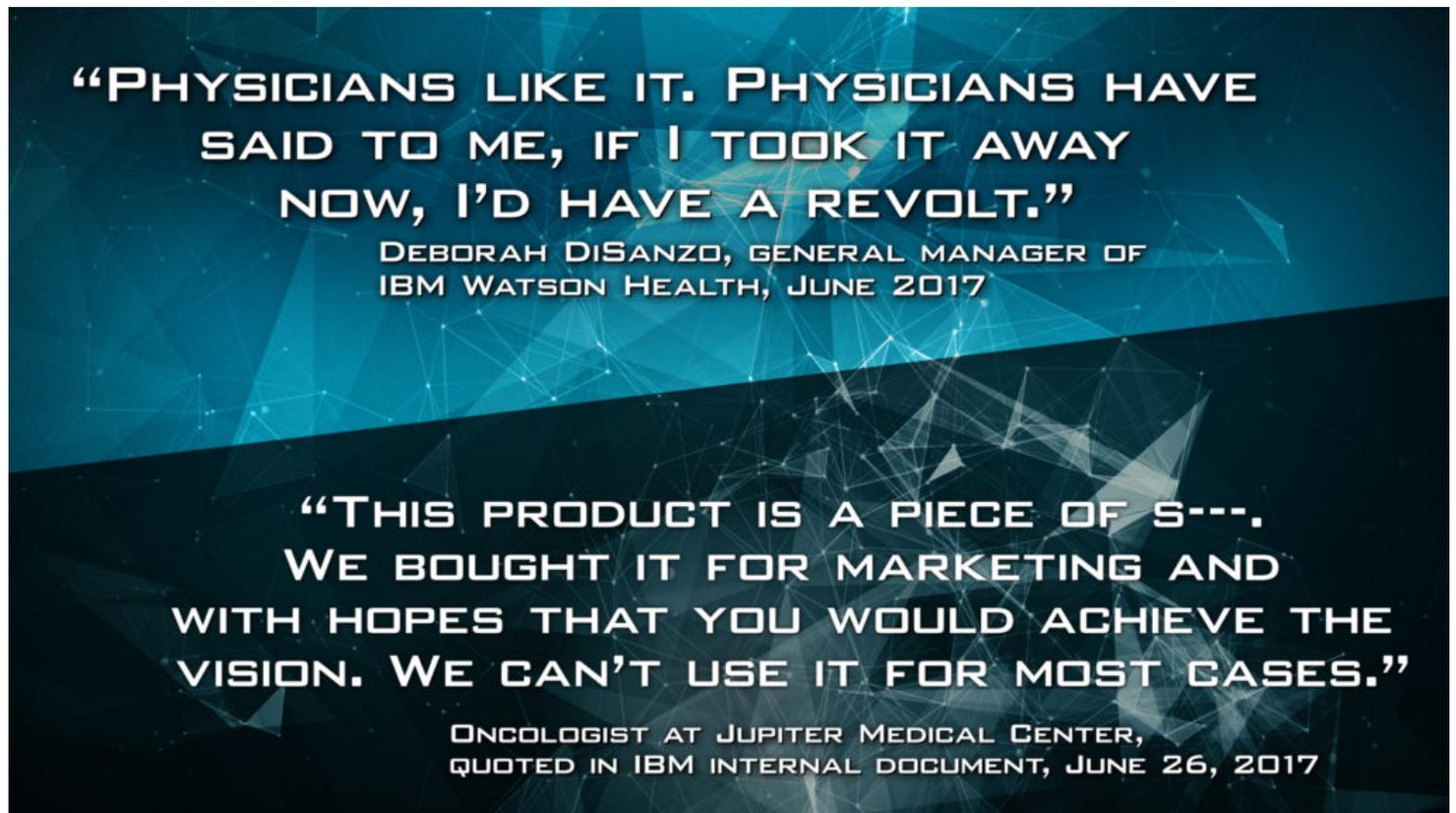
[Related:](#) <sup>1</sup>

[IBM pitched its Watson supercomputer as a revolution in cancer care. It’s nowhere close.](#) <sup>1</sup>

“Physicians like it. Physicians have said to me, if I took it away now, I’d have a revolt,” Deborah DiSanzo, general manager of IBM Watson Health, told STAT in a June 2017 interview.

As recently as two months ago, John Kelly, the senior vice president for IBM’s cognitive solutions division, which includes Watson, said at an [IBM event](#) <sup>2</sup> that Watson “has ingested all of the Memorial Sloan data, historic patients and results.” And at another [event](#) <sup>3</sup> in April, he said Watson for Oncology is “going fabulously.”

“It’s critical that everyone involved know what the data was that it was trained on,” Kelly added, echoing September 2017 [comments](#) <sup>4</sup> from IBM CEO Ginni Rometty, who said, “You must be transparent about it because that matters in these decisions.”



STAT first published [an investigation](#)<sup>1</sup> about problems with Watson for Oncology last September, reporting that it was not living up to the company's expectations, and that it was generating complaints from doctors around the world that its recommendations often weren't appropriate for patients in their countries. As a result, IBM sullied its reputation in the burgeoning global market for tools using artificial intelligence to improve cancer care, a sector that is a top priority for IBM and potentially worth billions of dollars.

But these new documents reveal that the problems were more serious and systemic and that IBM executives knew that the product was generating inaccurate recommendations that were at odds with national treatment guidelines — although there's no mention that patients were actually harmed. The documents also state that studies IBM conducted on the software, whose findings were touted as evidence of the system's usefulness, were designed to generate favorable results.

The documents were presented by Dr. Andrew Norden, an oncologist and deputy health chief, before he left IBM Watson Health last August. They show that, while IBM executives were talking up the product publicly, they were hearing harsh comments from customers. Even doctors at hospitals helping to promote the product were telling IBM executives privately that it was not useful in treating patients.

“This product is a piece of s—,” one doctor at Jupiter Hospital in Florida told IBM executives, according to the documents. “We [bought it for marketing](#)<sup>6</sup> and with hopes that you would achieve the vision. We can't use it for most cases.”

[Related:](#)<sup>6</sup>

[A new advertising tack for hospitals: IBM's Watson supercomputer is in the house](#)<sup>6</sup>

IBM defended its Watson for Oncology software, saying in a statement to STAT: “We have learned and improved Watson Health based on continuous feedback from clients, new scientific evidence and new cancers and treatment alternatives. This includes 11 software releases for even better functionality during the past year, including national guidelines for cancers ranging from colon to liver cancer.”

It said Watson for Oncology is trained to help treat 13 cancers, and is used by 230 hospitals worldwide.

Caitlin Hool, a spokeswoman for Memorial Sloan Kettering, said in a statement that the internal documents critical of the system's training and performance reflect “the robust nature of the process” of building and deploying the product in clinical care. Hool said the cancer center is continuously working with IBM to improve the accuracy and breadth of the system's recommendations.

“Patient safety is paramount,” Hool said. “While Watson for Oncology provides safe treatment options, treatment decisions ultimately require the involvement and clinical judgement of the treating physician. ... No technology can replace a doctor and his or her knowledge about their individual patient. To that point, the tool is also not equivalent to the cancer care delivered at MSK.”

Norden declined to comment, emailing STAT, “As you are aware, I am no longer employed by IBM Watson Health, and as such I am unable to discuss IBM or its business.”

Jupiter Medical Center did not respond to requests for comment.



9

IBM chief health officer Dr. Kyu Rhee, left, speaks with Dr. Andrew Norden, then Watson Health deputy health chief, at the American Society of Clinical Oncology annual meeting in 2017. Heather Stone for STAT

IBM struck a deal with Memorial Sloan Kettering to train Watson to help treat cancer patients in 2012, and began selling the product in Asia a few years later even though it had only been trained on a handful of cancers.

The presentations by Norden call out an array of purported flaws in the training methods, including the small number of cases used, which it says was “determined without statistical input,” and differences between MSK’s treatments and standard guidelines. They also note that Watson for Oncology was slow to adapt to new research findings and changing treatment guidelines.

The July 27 document stated the training and effectiveness of the product was undermined by the “inadequacy of the training cases.” It said one or two doctors trained the system to give treatment recommendations for each type of cancer, and that the cases were “synthetic,” meaning that they were devised by MSK doctors and were not real patients.

The synthetic cases were compiled by the doctors and IBM engineers to expose Watson for Oncology to clinical scenarios, as opposed to the actual records of patients who were treated at the hospital. That meant that Watson’s recommendations were driven by the doctors’ own treatment preferences — not a machine learning analysis of real patient cases.

[Related:](#) <sup>10</sup>

The training methods, and Watson's advice, led to loud complaints from doctors, according to the presentation, contributing to client dissatisfaction and physician concern. It also stated that Memorial Sloan Kettering's recommendations deviated from guidelines published by the National Comprehensive Cancer Network, a frequently referenced source of treatment recommendations, and in some instances reflected an "unconventional interpretation of evidence."

The presentation cited as an example Watson's recommendation that a 65-year-old man with newly diagnosed lung cancer and evidence of severe bleeding be given combination chemotherapy and a drug called bevacizumab. A "black box" warning for the drug, sold under the brand name Avastin, cautions it can lead to "severe or fatal hemorrhage" and shouldn't be administered to patients experiencing serious bleeding.

"Oy vey," Dr. Eric Topol, director of the Scripps Translational Science Institute, told STAT upon hearing about the error. "Any time you have an algorithm that makes a recommendation that's dangerous, that's extremely worrisome. I mean, the whole idea is that algorithms are supposed to improve safety and quality."

In its statement, Memorial Sloan Kettering said it believes the lung cancer recommendation cited in the presentation was part of IBM's system testing, and was not given to a real patient. "This is an important distinction and underscores the importance of testing and the fact that the tool is intended to supplement – not replace — the clinical judgement of the treating physician," the statement said.

The cancer center also said that in 2014, when Watson for Oncology was still in development, historical patient cases were initially used to train the system. But IBM determined that the synthetic cases — designed to be representative of cohorts of actual MSK patients — were better suited for the development of Watson for Oncology.

"The speed at which standards of care have changed require a more dynamic approach than historical data can provide because historical cases do not necessarily reflect the newest standards of care," MSK's statement said. "Synthetic cases also allow for diverse treatment options to be included in Watson for Oncology's recommendations, rather than a more narrow focus of how individual patients were treated at MSK."

But [product information](#)<sup>11</sup> posted on the IBM website and dated Feb. 15, 2017, implies that Watson continues to be trained with real patient data. It states that Watson for Oncology "analyzes patient data against thousands of historical cases and insights gleaned from thousands of Memorial Sloan Kettering MD and analyst hours."

The July 2017 presentation shows the number of cases used to train Watson as of that date for eight different cancers; they range from 635 cases for lung cancer to 106 for ovarian.

Experts in artificial intelligence told STAT that IBM's portrayal of the training, and the number of doctors and patients involved, raises questions about whether it's being transparent with users about the source and value of Watson's recommendations.

“The thing which is a bit misleading is that everybody’s led to believe that this is the consensus of the entire brain trust of Sloan Kettering,” said Nigam Shah, associate professor of medicine and biomedical data science at Stanford. “But in fact it’s the consensus of ... a small subset of the entire brain trust.”

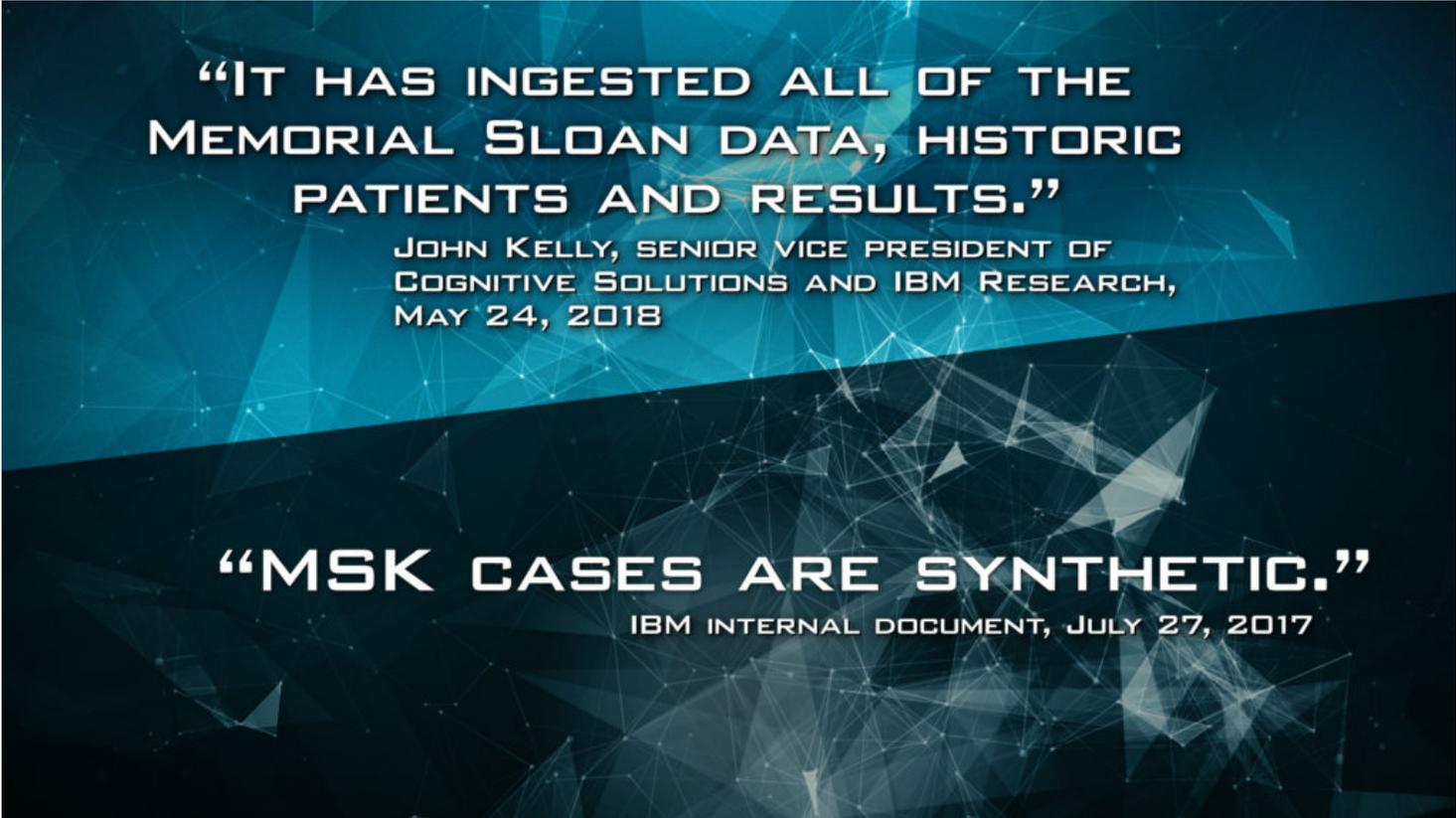
“They should be called out on this,” Shah added. “I would bet this is a calculated risk they took. ... They’re kind of messing with people, but it’s within the marketing spin that is increasingly allowed these days, let me put it that way. But not everybody can spot it, so it’s not honest.”

Some experts said it’s possible that hypothetical patient data would do a good job of training the system — if it was representative of real patient data.

“I would certainly want to see some validation to whether the synthetic data is representative of anything that would make sense,” said Dr. Jonathan Chen, an assistant professor at Stanford’s Center for Biomedical Informatic Research.

Jana Eggers, CEO of Nara Logics, an artificial intelligence company, said Watson’s use of synthetic data made clear that this software was not making use of the “big data” that exists in health care — troves of information about individuals that are too complex or burdensome for humans to navigate.

“They’re making up personas of cancer patients, basically,” Eggers said. “Why do you do that when you have the real people?”



**“IT HAS INGESTED ALL OF THE  
MEMORIAL SLOAN DATA, HISTORIC  
PATIENTS AND RESULTS.”**

JOHN KELLY, SENIOR VICE PRESIDENT OF  
COGNITIVE SOLUTIONS AND IBM RESEARCH,  
MAY 24, 2018

**“MSK CASES ARE SYNTHETIC.”**

IBM INTERNAL DOCUMENT, JULY 27, 2017

The internal documents also raise questions about the validity of the studies conducted by IBM that the company used to demonstrate the value of Watson for Oncology to doctors around the world.

In recent years, IBM and its clinical partners have published [multiple studies](#)<sup>13</sup> demonstrating that Watson would achieve a high level of “concordance” with the treatment recommendations of oncologists. That is theoretically valuable because it shows that Watson could generate recommendations within seconds that doctors would otherwise spend many hours or days developing.

But one of the presentations states that the concordance studies had been “designed in a way that makes negative findings unlikely.” It noted that sophisticated users of the system would demand “robust prospective evidence” of compliance with guidelines, cost savings, and improvement on quality metrics.

“The system as currently designed is unlikely to impact any of the above,” Norden’s June 26, 2017, presentation said.

It cautioned, however, about the risks of conducting such a study at a time when IBM was already selling the system around the world: “In my view, conducting a rigorous study now is a very high-risk endeavor that could be embarrassing at best and have serious adverse business consequences at worst.”

Indeed, a company working with IBM to implement Watson for Oncology at hospitals in the Netherlands told STAT in June that they aren’t interested in “concordance” or whether Watson recommends a treatment that is the same as treatment guidelines.

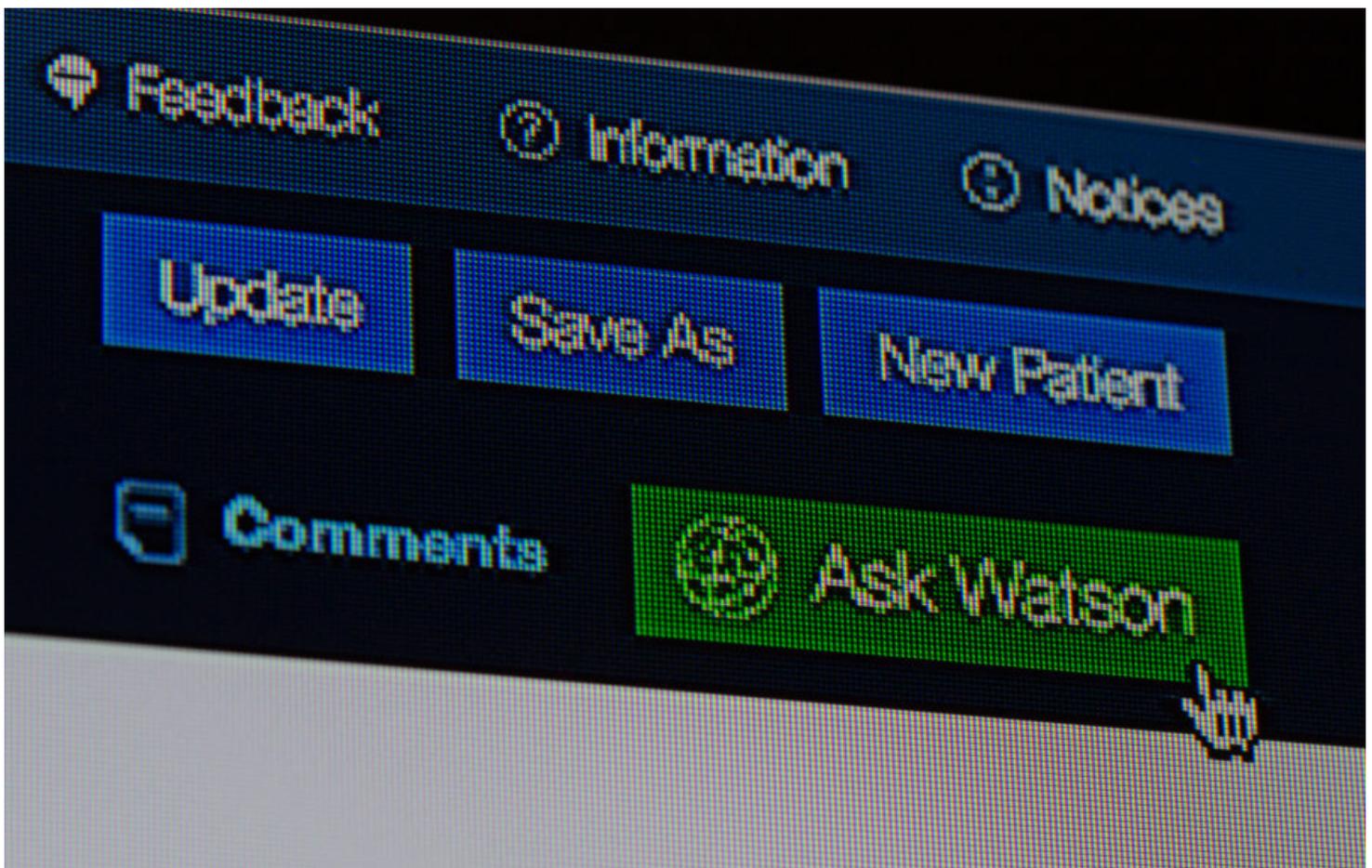
“What we are interested in is whether the system will challenge us in our thinking,” said Vincent The, head of strategy, research, and development for MRDM, the Dutch company.

He said MRDM is working with IBM to improve Watson for Oncology. “It seems to be maturing quite a bit,” The said. “However, at the current state, it’s not at the level we need it to be.”

Many users of Watson for Oncology, as well as IBM’s own employees, were raising concerns about the system’s performance last year. Norden’s presentation stated that a number of employees, including developers and oncologists, had “confided” in him that the product was “very limited.”

Norden’s presentation offered a couple of solutions for improving the system: the number of patient cases could be increased to reflect “MSK practice patterns,” with minimum numbers to be set based on statistician input. Or, he said, standard treatment guidelines could be used as the basis of Watson’s recommendations, customized for different locations, and Memorial Sloan Kettering could identify its own institutional approaches within the system.

It is not clear whether any of those suggestions was implemented.



14

Heather Stone for STAT

Norden left IBM in August 2017 to take a job at Cota Inc., another company seeking to analyze cancer data that had already struck up a partnership with IBM.

Since June 2017, the two companies have been working on a joint product that would combine Cota's technology with Watson. Cota can compare individual cancer patients to a private database of how other patients were treated to help a doctor determine the best method of treatment.

Norden's presentation suggested that the work with Cota, a company founded by a doctor at Hackensack Meridian Health in New Jersey, could help expose Watson for Oncology to the real-world evidence needed to improve its recommendations.

Doctors at Hackensack Meridian have meanwhile completed a pilot project testing the joint product.

"It went well," Norden said of the project in a July 10 interview with STAT, explaining that it was intended to create a tool physicians could use at the point of care to highlight the types of treatments that had worked best for specific types of patients.

"We're taking the pilot and scaling it up," Norden said, declining to elaborate.

Memorial Sloan Kettering had also noticed the promise of Cota's product: In November 2017, it struck a deal to have Cota help analyze the medical center's historical patient records. As part of the deal, MSK received equity in Cota.

In an emailed response to STAT questions about the reason for the partnership, Hool, the MSK spokeswoman, wrote: “There are important insights to be gained from the information in patients’ medical records, but these insights are often inaccessible due to unstructured, disjointed data within the medical record. Cota has an innovative platform and approach for taking patient records and turning them into data sets that can be analyzed for insights into diagnosis, care pathways, and outcomes.”

## About the Authors



[Casey Ross](#)<sup>15</sup>

National Correspondent

Casey covers the companies disrupting biopharma and health care.

[casey.ross@statnews.com](mailto:casey.ross@statnews.com)<sup>16</sup>

[@caseymross](#)<sup>17</sup>



[Ike Swetlitz](#)<sup>18</sup>

Washington Correspondent

Ike is a Washington correspondent, reporting at the intersection of life science and national politics.

[ike.swetlitz@statnews.com](mailto:ike.swetlitz@statnews.com)<sup>19</sup>

[@ikeswetlitz](#)<sup>20</sup>

## Tags

## Links

1. <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>
2. <https://www.govexec.com/feature/ibm-think-gov-2018-livestream/>
3. <https://www.youtube.com/watch?v=b04gEbHuFDg>
4. <https://www.bloomberg.com/news/features/2017-09-20/ginni-rometty-on-artificial-intelligence>
5. [https://www.statnews.com/wp-content/uploads/2018/07/WATSON\\_QUOTE\\_01.jpg](https://www.statnews.com/wp-content/uploads/2018/07/WATSON_QUOTE_01.jpg)
6. <https://www.statnews.com/2017/09/06/watson-ibm-marketing/>
7. <https://www.statnews.com/signup/>
8. <https://www.statnews.com/privacy/>
9. <https://www.statnews.com/wp-content/uploads/2018/07/WATSON0011.jpg>
10. <https://www.statnews.com/2018/06/11/ibm-watson-health-problems-layoffs/>
11. [https://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep\\_oc/1/897/ENUS5725-W51/index.html&request\\_locale=en](https://www-01.ibm.com/common/ssi/ShowDoc.wss?docURL=/common/ssi/rep_oc/1/897/ENUS5725-W51/index.html&request_locale=en)
12. [https://www.statnews.com/wp-content/uploads/2018/07/WATSON\\_QUOTE\\_02.jpg](https://www.statnews.com/wp-content/uploads/2018/07/WATSON_QUOTE_02.jpg)
13. <https://www-03.ibm.com/press/us/en/pressrelease/52502.wss>
14. <https://www.statnews.com/wp-content/uploads/2018/07/WATSON0005.jpg>
15. <https://www.statnews.com/staff/casey-ross/>
16. <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/mailto:casey.ross@statnews.com>
17. <https://twitter.com/caseymross>

18. <https://www.statnews.com/staff/ike-swetlitz/>
19. <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/mailto:ike.swetlitz@statnews.com>
20. <https://twitter.com/ikeswetlitz>
21. <https://www.statnews.com/tag/cancer/>
22. <https://www.statnews.com/tag/medical-technology/>
23. <https://www.statnews.com/tag/stat-plus/>