difficulties. But several researchers suggest that the problems may be due to widespread media coverage of this year's severe US flu season, including the declaration of a public-health emergency by New York state last month. The press reports may have triggered many flu-related searches by people who were not ill. Few doubt that Google Flu will bounce back after its models are refined, however.

"You need to be constantly adapting these models, they don't work in a vacuum," says John Brownstein, an epidemiologist at Harvard Medical School in Boston, Massachusetts. "You need to recalibrate them every year."

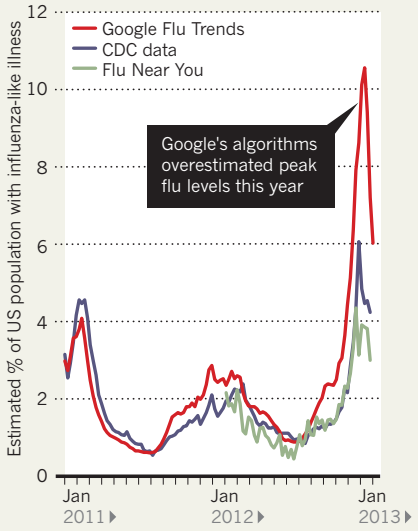
Brownstein is one of many researchers trying to harness the power of the web to establish sentinel networks made up not of physicians, but of ordinary citizens who volunteer to report when they or someone in their family are experiencing symptoms of ILI. 'Flu Near You', a system run by the HealthMap initiative co-founded by Brownstein at Boston Children's Hospital, was launched in 2011 and now has 46,000 participants, covering 70,000 people.

Similar systems are springing up in Europe. For example, GrippeNet.fr, run by French researchers in collaboration with national health authorities, has attracted more than 5,500 participants since its creation a year ago, with 60–90 people joining each week.

Lyn Finelli, head of the CDC's Influenza Surveillance and Outbreak Response Team, feels that such crowdsourcing techniques hold great promise, especially because the questionnaires are based on clinical definitions of ILI and so yield very clean data. And both Flu Near You and GrippeNet.fr have a representative age distribution of participants. The CDC has worked with Flu Near You on its development, and Finelli herself has signed up: "I submit my family's data every week," she says.

FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



SOURCES: GOOGLE FLU TRENDS (WWW.GOOGLE.ORG/FLUTRENDS); CDC; FLU NEAR YOU

Other researchers are turning to what is probably the largest publicly accessible alternative trove of social-media data: Twitter. Several groups have published work suggesting that models of flu-related tweets can be closely fitted to past official ILI data, and various services, such as MappyHealth and Sickweather, are testing whether real-time analyses of tweets can reliably assess levels of flu.

But Finelli is sceptical. "The Twitter analyses have much less promise" than Google Flu or Flu Near You, she says, arguing that Twitter's signal-to-noise ratio is very low, and that the most active Twitter users are young adults and so are not representative of the general public.

Michael Paul, a computer scientist at Johns Hopkins University in Baltimore, Maryland, disagrees. He is part of a team that is developing Twitter-based disease monitoring, and says that Google search-term data probably have just as much noise. And although Internet-based surveys may boast less noise, their smaller size means that they may be prone to sampling errors. "I suspect that passive monitoring of social media will always yield more data than systems that rely on people to actively respond to surveys, like Flu Near You," Paul says.

To reduce the noise, the Johns Hopkins team has recently analysed a subset of a few thousand flu-related tweets, looking for patterns indicating which tweets showed that the tweeter was actually ill rather than simply, say, pointing to news articles about flu. They then used this information to retrain their models to weed out irrelevant flu-related tweets. Paul says that a paper in press will show that this greatly improves their results.

Already, web data mining and crowdsourced tracking systems are becoming a part of the flu-surveillance landscape. "I'm in charge of flu surveillance in the United States and I look at Google Flu Trends and Flu Near You all the time, in addition to looking at US-supported surveillance systems," says Finelli. "I want to see what's happening and if there is something that we are missing, or whether there is a signal represented somewhat differently in one of these other systems that I could learn from." ■

MEDICINE

Data barriers limit genetic diagnosis

Tools for data-sharing promise to improve chances of connecting mutations with symptoms of rare diseases.

BY ERIKA CHECK HAYDEN

For the first five months of Harrison Harkins' life, doctors had little idea about what was causing his spinal malformation and inability to gain weight. But in November 2011, Matthew Bainbridge, a computational biologist at Baylor College of Medicine in Houston, Texas, found a clue. After analysing genetic data from Harrison and his parents, Bainbridge discovered that the child had an abnormal version of a gene called *ASXL3*.

But Bainbridge had no easy access to records of other children with *ASXL3* mutations, and could not be sure that this mutation was the culprit. So he did what many scientists do: he networked. A Dutch team put Bainbridge in touch with German researchers who were treating another boy with an *ASXL3* mutation — and symptoms similar to Harrison's. After finding two further cases in an internal Baylor database, Bainbridge felt that the connection was concrete. He describes the syndrome seen in all four children, and probably caused by *ASXL3* mutations, in a paper published on 5 February (M. N. Bainbridge *et al. Genome Med.* **5,** 11; 2013).

Researchers are using new tools to increase the pace of discoveries such as Bainbridge's. Efforts to connect sequences with symptoms — or in genetic parlance, genotype with phenotype — have taken on increased urgency as clinical sequencing gains traction and funders put more money towards rare diseases. Researchers are planning to address the barriers to data sharing at a workshop in April, after the first International Rare Diseases Research Consortium Conference in Dublin. "There is a very positive feeling in the community that things are changing for the better," says Peter Robinson, a computational biologist at the Charity University Hospital in Berlin.

Thousands of people have had their genomes sequenced, but a reluctance to surrender ownership of the valuable data, along with the privacy concerns of researchers and families (see 'Families find solace in

A RARE CONNECTION

Families find solace in sequencing

Medical researchers versed in the power of genetic sequencing may be surprised by how unaware of it — or doubtful about its benefits — the rest of the world remains. Tim Harkins, who develops sequencing technologies at Life Technologies in Carlsbad, California, found this out when he suggested that doctors sequence part of his ailing son's genome to help with the child's diagnosis. They "didn't have a clue what I was talking about", says Harkins.

So he approached geneticists at Baylor College of Medicine in Houston, Texas, who discovered a mutation in Harrison's *ASXL3* gene. The diagnosis "was an invaluable source of comfort", says Harkins, not least because he could tell his older son, now aged 13, that the mutation was spontaneous and not heritable. "I was able to tell him, 'You have a better chance of being struck by lightning than having a baby like Harrison'," says Harkins.

Despite the rapidly dropping cost of sequencing, Harkins notes, there is still a perception that it is too expensive for clinical use and does not deliver meaningful results. But his own experience gives the lie to that sentiment, he says. "Knowing the diagnosis can prevent a lot of needless surgeries and treatments and it eliminates the guessing, the 'Why don't we try this or that?'."

BRIAN MCGUCKIN

Harrison Harkins died one month after his final diagnosis.

The unease runs even deeper in Germany. Maria and Christian-Alexander Neuling, who live in Hamburg, found some comfort after a genetic diagnosis showed that their son, Ferdinand, also had an *ASXL3* mutation. The diagnosis helped to reassure the Neulings that they were doing everything they could for Ferdinand, who is now four; it also enabled them to connect with another family in the United States whose child has the same condition.

But they worry that genetic sequencing is viewed too sceptically in Germany, where a doctor's orders and mandatory counselling are required for any genetic test. "The discussion here is pretty one-sided; there is a strong lobby opposing genetic sequencing," says Christian-Alexander Neuling. Neuling and his wife hope that their story will assuage some of the fears of other Germans, who worry that sequencing will lead to eugenics and 'designer babies'. "A parent will go from hell to heaven and back to find something that will help their sick child," he says. **E.C.H.**

sequencing') often keep scientists from comparing findings. Many data are also off-limits because they are held by private diagnostic companies. "It's a big conundrum for labs that are doing sequencing for diagnostic services," says Michael Bamshad, chief of paediatric genetic medicine at the University of Washington in Seattle. "If they find a variant in a gene, how do they know the variant is causal?"

Patients with rare, difficult-to-diagnose disorders stand to gain the most from increased data sharing. Scientists have found the genetic roots of fewer than half of the 7,000 known rare heritable diseases, but a diagnosis can give parents an idea of a child's outlook, and give researchers a target for drug development.

Several groups are trying to build richer databases and get them to communicate. In November, for instance, the US National Center for Biotechnology Information in Bethesda, Maryland, set up a database called ClinVar, which pools information from dozens of other databases, and allows labs to deposit data on mutations seen in individual patients (see *Nature* **491,** 171; 2012).

Still, says Anthony Brookes, a geneticist at the University of Leicester, UK, many diagnostic labs are unable to share information with databases such as ClinVar, either because they do not have the time or the expertise in depositing data, or because they are afraid that they might compromise patient security and their own livelihoods. "It's not their role to put data out there for researchers to play with," he says.

Brookes is trying to address this problem with a tool called Cafe Variome, which he describes as more of a "shop window" than a database. Labs submit information about what data they have to Cafe Variome. Users can then browse the website to see what data exist, and, if interested, can follow up with the relevant labs. That allows the labs to control who sees their data, and to be credited when it is used. They are "much more comfortable sharing if they know the data are only being accessed by other diagnostic labs", says Brookes.

Another problem is that even if database owners are willing to share data, they lack a common language for describing phenotypes, says Robinson. He is working on ways to standardize phenotype definitions for large-scale analysis.

For researchers such as Bainbridge, the tools can't come quickly enough. His team's final diagnosis came just a month before Harrison died last March, at the age of 9 months. "If you spent 15 minutes with the parents of any of these children, you would know that everyone should be doing this," says Bainbridge. "This is going to help a lot of people at really low cost." ■

MORE ONLINE

CARL BUELL

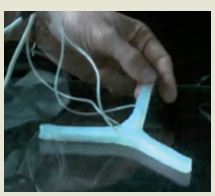
TOP STORY

Data reveal common ancestor of placental mammals go.nature.com/9ezari

MORE NEWS

- Plant experiments renew controversy on extra-genome inheritance go.nature.com/ukvsc3
- Small molecule triggers cancer-cell suicide in mice go.nature.com/riagfg
- How to turn living cells into computers go.nature.com/1v6bzp

VIDEO



Controlled explosions make rubbery robot jump go.nature.com/l8395v