

AN AUDIENCE WITH...

Charles Hugh-Jones

Drug developers can identify new opportunities and get insight into their failures when they mine their in-house clinical trial data. But to gain maximum insight, companies would do well to pool and share their data sets, argues Charles Hugh-Jones, Chief Medical Officer in North America at Sanofi. To this end, he and his collaborators from the CEO Roundtable on Cancer are preparing to launch [Project Data Sphere](#), a non-profit platform designed to share the comparator arm data from dozens of cancer clinical trials. Hugh-Jones talks to **Asher Mullard** about the need for, and the promise of, data sharing.

Q *How did Project Data Sphere get started?*

In 2001 President Bush started the CEO Roundtable on Cancer to look into what we could do to speed cancer research. 7.6 million people worldwide die from cancer every year, and despite huge improvements in disease understanding we haven't developed interventions that substantially shift this overall mortality rate.

The question for us was what could we — different pharmaceutical companies and academic institutions, with input from the US Food and Drug Administration (FDA), the Institute of Medicine and others — do as a group that each of us couldn't do on our own? Data sharing interested us, so two and a half years ago we started setting up a platform with SAS that would share the comparator arm data from cancer clinical trials.

So far we've got data from ten Phase III trials, from seven companies and three academic organizations. But we'd like to get data from 50 trials by the end of next year. SAS has provided analytical tools and we have ensured the platform is broadly accessible. We have plans to set up discussion boards around each area, around research interests and disease areas, and so on. We want the community to contribute to an ecosystem and establish a dialogue.

Q *Why are you only sharing comparator arm data?*

We tried to do what was achievable rather than what was ideal. But there are all kinds of useful information in there. What you get is the full comparator arm, of which there are several hundred fields of data for several hundred patients in each data set. You get the full protocol, case report forms, and more.

Q *What do you think will come out of Project Data Sphere?*

What we are providing is the raw ingredients: clinical data and analytics tools. But our whole goal is to see what happens when we share the data. We want to source the crowd to see what work they want to do. Can we explore novel statistical techniques?

Can we do real-world versus comparator trial comparisons? Can we use these data as a quasi-comparator arm in future studies? Can these data be used to develop natural history progression models?

Another goal is to make different organizations — academic and commercial — understand what can be done with these types of platforms, and show that there is a reasonable way of open access sharing for certain data.

But every time we present this project, someone comes up to us afterwards and suggests a different and interesting way of using the data. That's exciting. When the first code for the internet was written, nobody would have dreamt that the internet would evolve to be what it is now. I'm much more excited by what we don't know can be done with these data, versus what we already know can be done with them.

Q *Are you developing data standards to get maximal value out of shared and pooled data?*

That's for other people to do. We are providing one step in the value chain, and what we want to see is other people feed things — such as data standards — back into Project Data Sphere.

But, an interesting conversation I had very early on in the project was with Becky Kush, CEO at the Clinical Data Interchange



CEO Roundtable on Cancer

Standards Consortium (CDISC). She said that one of the problems in terms of developing disease-specific data standards was in getting copies of protocols and case report forms and data sets from four or five different companies so that she could compare them. And this is just one example of what this type of sharing can do.

Q *Will you include genetic data?*

The historical data sets providers are including first are from 2–5 years ago, when it wasn't routine to include large genomic panels. We are now thinking about crossing some of those barriers.

But, there is always a trade-off between access and data. If you've done whole-exome sequencing, you are not going to be able to pool it in an open access system because of the privacy issues. If you have data on single mutations, by contrast, you are probably OK in terms of privacy. We would probably have to end up with some kind of access 'tiering' if we want to include genomic data.

Q *Other groups are also working on developing clinical trial data sharing networks, including ITN Trial Share and the Pooled Resource Open-access ALS Clinical Trials (PRO-ACT). As a whole field, what stage are you at?*

We are still in a big dark room, feeling out the boundaries of the wall. We still need to figure out what is acceptable in terms of privacy, access, analytics and types of data. I certainly support multiple initiatives to explore sharing, rather than just a few. It is very likely that there won't be a one-size-fits-all approach. Sharing is complex and diversity will support innovation, which is good for patients.