Learning from the pioneers:

Lessons about data platforms drawn from the WorldWide Antimalarial Resistance Network (WWARN) experience

Researchers are increasingly being encouraged to share individual patient data from clinical trials, but there's remarkably little experience about how that can be done equitably, ethically and efficiently, especially for diseases where most research takes place in low and middle income countries. One pioneer in this area is the Worldwide Antimalarial Resistance Network (WWARN), conceived almost a decade ago by malaria researchers from across the globe to bring together clinical, in vitro, molecular, pharmacological and (later) medicine quality data. It was hoped these data, analysed together, would allow for the efficient tracking of drug-resistant malaria.

WWARN facts and figures (at October 2016)

Individual patients included in database:	135,000
Clinical trials included in database:	186
% of all published trials of artemisinin combination therapies:	80%
Molecular studies included in database	103
Number of data contributors and other collaborators	268
Published individual patient meta-analyses using WWARN-standardised data	17
Principal funder:	Bill and Melinda Gates Foundation
Additional support:	ExxonMobil Foundation
	Ministry of Foreign Affairs, France
	European & Developing Countries Clinical Trials Partnership
	UK Medical Research Council, UK Department for International Development

Researchers and those that fund them, together with the journals that publish research results, are all moving towards a more "open" model of science, of which data sharing is an important part. To contribute to discussions about how best to share data, the Wellcome Trust -- on behalf of the Public Health Research Data Forum -- commissioned a study to capture the learning provided by the experience of WWARN, which is both rich and still rare. WWARN founders, staff and collaborators were generous in collaborating with the study, which was carried out by independent researchers. The study, which was designed as a learning exercise rather than a formal evaluation, was based on a comprehensive records review, in-depth interviews with 47 people involved with WWARN (including some who have chosen not to contribute data to the platform), and a witness seminar which yielded sometimes divergent views about the genesis and development of the network.

The aims of WWARN shifted over time but there's no doubt that the collaboration has contributed to a better understanding of malaria treatment efficacy. Pooled analyses based on data compiled and standardised by WWARN have informed changes to international guidelines on antimalarial treatment and dosage (see box, below). The collaboration, housed at Oxford University since 2009, is now under the umbrella of the Infectious Disease Data Observatory (IDDO). Supported by the Wellcome Trust, Médecins Sans Frontières, the WHO-affiliated Special Programme for Research and Training in Tropical Diseases (TDR), Drugs for Neglected Diseases Initiative and Foundation for Innovative

New Diagnostics, IDDO is adapting the platform for Ebola, visceral leishmaniasis and other diseases.

This document summarises seven key lessons around the factors that make sharing individual patient and pathogen data feasible and useful. We hope that WWARN's experience will help inform the efforts of other researchers and, particularly, those of funders, policy-makers, companies and other communities that support and benefit from scientific enquiry, as they work to share knowledge to improve health.

Lesson 1: Data sharing platforms work well when the motivations of platform funders, developers, contributors and users are considered at the outset, and the incentives driving each of them are aligned.

WWARN initially hoped to provide national policy-makers with real-time information about the geographic spread of drug-resistant malaria. This goal was not realised because incentives were misaligned. There was little demand for global analysis from national policy makers. More importantly, clinical trialists in endemic countries (who still advance through publication in peer reviewed journals) did not want to share data with other scientists before they had published their own results.

The platform took off after WWARN switched its focus to pooled analyses. WWARN now requests data from principal investigators conducting efficacy trials of antimalarial drugs to answer specific research questions. Data contributors are invited to participate in analysis and paper writing, and are credited appropriately on resulting publications. This motivated researchers to contribute their data to the curated resource, and has produced world-class science: new methods have been developed and pooled analyses have led to changes in international malaria treatment guidelines. Finally, potential users, including drug developers and the WHO, are approaching the platform with specific questions.

Lesson 2: Both disease experts and data scientists are core to the design of a successful data sharing platform with public health aims

The malaria specialists who conceived of WWARN wished to minimise barriers to sharing and maximise flexibility of potential outputs. While they put in countless unpaid hours debating which information was most useful and developing standards, they resisted pre-defining the end uses of the database. Information scientists, who know how to develop shared resources to achieve network effects, pressed for greater clarity of purpose at the outset. In retrospect, defining the core purpose of the database more clearly at the start -- a task for disease experts -- would have resulted in a more efficient development process.

Lesson 3: Data curation is expensive, but essential if shared data are to be useful

WWARN invested a great deal of time and effort developing the standards and the tools to allow datasets to be standardised and combined across time and location. It is these investments that make the shared data useful. The value of a well-curated dataset grows over time but the costs are front-loaded; investors should not expect a quick return. However the cost of curation tools developed for one disease platform may be apportioned across future platforms, because instead of developing new systems from scratch, many existing structures can be adapted.

Lesson 4: Data sharing platforms should be underpinned by clear, equitable governance structures that can evolve in line with changing community norms.

Less visible than the curation tools, but just as important, are the procedures WWARN has developed to govern contribution and use of data shared through the platform. From the start, data use agreements were clearly worded and not overly legalistic. Developed in uncharted territory in the face of widespread scepticism, the early terms of submission required permission from investigators for every use of shared data -- a huge administrative burden for the platform. Community confidence and norms supporting data sharing have since grown, and with them the possibility for more open models of sharing. Terms which default to greater sharing while allowing contributors to exercise more control if desired may help balance trust-building with the flexibility to evolve along with the data sharing zeitgeist.

New partnerships lead to rapid, policy-relevant analyses

More than a decade ago, researchers from the Liverpool School of Tropical Medicine collaborated with the WHO-housed Special Programme for Research and Training in Tropical Diseases (TDR) to show that weight-for-age translations were inaccurate in some parts of the world, meaning that children may be getting inappropriate doses of medication. Their re-calculated rates were used by pharmaceutical firm Sanofi in formulating fixed dose combination of the antimalarial artesunate - amodiaquine.

In late 2013, WWARN published an analysis based on data from 7,072 patients, pooled across 26 studies. They showed that young children taking dihydroartemisinin-piperaquine seemed to be getting a lower intake dose than adults, and were at higher risk for recrudescence. This suggested that children may need higher doses, but it wasn't known if higher doses would be safe. This evidence was independently reviewed by WHO, which changed its dosing recommendations for children taking the drug. The Liverpool group quickly switched the focus of a planned dosing study, looking instead at the efficacy and risk of cardiotoxicity of a higher dose of DHA-piperaquine in children. In short, a hypothesis derived from pooled analysis of well-curated, shared data led rapidly to a targeted clinical trial. An urgent question was answered, and policy quickly changed. The result should be fewer treatment failures and more healthy children.

Lesson 5: Institutional arrangements have important implications for data sharing; these should be considered with care at the outset of a data sharing venture.

Initiated by a small group of relatively well-resourced malaria specialists in consultation with senior researchers from many malaria-endemic countries, WWARN was conceived as a loose network of professional researchers. For practical reasons, driven in part by the needs of funders, it was then embedded within an academic department at Oxford University. This arrangement has had important consequences. Staff may have career expectations which are not advanced by their functional roles, particularly if they are subject to traditional publication-based measures of academic productivity. In addition, in a culture where "ownership" of data is still contested, siting a data sharing platform in a northern academic institution may create the perception of a "data grab" and limit the willingness of key partners, including endemic country researchers, to participate fully.

Lesson 6: More interaction with national authorities and participation of endemic country researchers in setting research questions may contribute to more locally relevant and actionable research results.

WWARN's current strength is now in producing high quality peer reviewed papers that make the best use of a large, multi-country database to yield learning of global significance. The platform has not yet found a ready audience among national policy makers, however.

A granular understanding of local data will become increasingly important as disease prevalence falls. Involvement of endemic country researchers in analysis and interpretation should thus also grow, but right now, endemic country researchers rarely use shared data resources. This is in part because those who have the skills to perform complex analyses are too senior to spend time on such tasks. Scientific collaborations, including data sharing networks, must be funded to work with partners to reinforce skills at more junior levels. Support should include financial and professional incentives that encourage endemic country researchers to conduct secondary analysis that answers questions raised by local policy makers.

Lesson 7: As data sharing platforms mature, institutional arrangements may shift

As lesson 2 suggests, scientists who specialise in disease areas are the critical drivers of platform development, and specialists in informatics design are essential to its conception. Disease specialists are also best placed to define early research questions, and will thus be key in the "proof-of-concept" phase, when the utility of the platform is demonstrated. Universities or other research bodies are thus ideal incubators for new disease platforms, but because of incentive structures and perceptions of conflict of interest, they may not be the best home for data platforms in the long term. Since there are so few models to draw on, it's impossible to recommend alternative models with confidence. However as data platforms mature, it is possible that day-to-day data management may be more cost-effectively handled by a neutral public health entity that employs a few specialist advisors and enforces standards and transparent governance structures developed and agreed by the broader scientific community. University-based disease specialists are likely to remain the most important users of a platform. They must continue to be involved in shaping its direction, without necessarily controlling it.

The full study, by Elizabeth Pisani and Stella Botchway is available at:
https://dx.doi.org/10.6084/m9.figshare.4476308.v1