Equitable data sharing: challenges and suggestions for ways forward [version 1; peer review: 2 approved with reservations]

Karen I. Barnes1-3*, Julio Arturo Canario4*, Sunil S. Vernekar5*, Shivaprasad S. Goudar5, Roberto Espinal4, Laura Merson6,7, Phaik Yeong Cheah7-9

1Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa
2MRC Collaborating Centre for Optimising Antimalarial Therapy, University of Cape Town, South Africa, University of Cape Town, Cape Town, South Africa
3WorldWide Antimalarial Resistance Network, Southern African regional hub, Department of Medicine, University of Cape Town, Cape Town, South Africa
4Centro Nacional de Investigaciones en Salud Materno Infantil Dr. Hugo Mendoza and Centro de Bioética Etikos, Santo Domingo, Dominican Republic
5Department of Physiology, KLE Academy of Higher Education and Research’s J.N. Medical College, Belgaum, India
6Infectious Diseases Data Observatory, University of Oxford, Oxford, UK
7Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK
8Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Please Select, 10400, Thailand
9The Ethox Centre, University of Oxford, Oxford, UK

* Equal contributors

Abstract
Data sharing is increasingly mandated by health research funders and publishers. In the context of data collected in low-resource settings, concerns have been raised regarding inequitable opportunities to engage in secondary use of data between researchers in well-resourced and resource-limited settings. In this context, we present three case studies and their issues related to equity: the multicountry Antenatal Corticosteroid Trial, health research in the Dominican Republic and the WorldWide Antimalarial Resistance Network. These case studies were discussed at the 2018 Global Forum for Bioethics in Research in South Africa, focusing on the theme “The ethics of data sharing and biobanking in health research”. The case studies provide concrete examples of real challenges such as lack of prior consent for data sharing, potential for misinterpretation of data by secondary users and limited capacity of researchers in low-resource settings to conduct secondary analyses. We conclude by suggesting ways forward. We stress the importance of capacity building and investments in data management and data science skills, and in data sharing platforms supporting poverty-related disease research. In addition, we recommend that there should be incentives to promote
data sharing and that research groups and institutions establish their own data sharing policies tailored to their context, data and community while pursuing harmonization with existing policies as much as possible. We also think that international guidelines on authorship criteria should be revisited. For new studies, researchers should obtain consent for sharing of participants’ data with secondary users. Lastly we recommend that community and stakeholder engagement be conducted to improve the consent process and identify what might be sensitive data to mitigate any potential harms to data subjects and their communities.

**Keywords**
Data sharing, low-resource settings, ethics, bioethics, equitable

This article is included in the GFBR: The ethics of data sharing and biobanking in health research collection.

**Corresponding author:** Phaik Yeong Cheah (phaikyeong@tropmedres.ac)

**Author roles:** Barnes KI: Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Canario JA: Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Vernekar SS: Conceptualization, Resources, Writing – Original Draft Preparation, Writing – Review & Editing; Goudar SS: Conceptualization, Resources, Writing – Original Draft Preparation; Espinal R: Conceptualization, Resources, Writing – Original Draft Preparation; Cheah PY: Conceptualization, Project Administration, Resources, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by the Wellcome Trust through funding to the Global Forum for Bioethics in Research (GFBR) and a Strategic award [096527] and the Wellcome Trust Thailand Major Overseas Programme [106698] which support PYC. GFBR funding is provided by the Wellcome Trust, the Bill & Melinda Gates Foundation [OPP1151904], the National Institutes of Health and the UK Medical Research Council. The WorldWide Antimalarial Resistance Network (WWARN) is funded by the Bill and Melinda Gates Foundation and ExxonMobile Foundation. The National Research Ethics Committees Survey was funded by the National Research Center on Child and Maternal Health Dr. Hugo Mendoza (CENISMI). We thank the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)'s Global Network for Women and Children's Health Research, USA for funding the Global Network, which includes seven low & middle income countries around the world including India. We thank the Women's and Children's Health Research Unit, KLE Acedemy of Higher Education and Research, Belagavi, Karnataka, India, the site which conducted the ACS trials.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2019 Barnes KI et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Barnes KI, Canario JA, Vernekar SS et al. Equitable data sharing: challenges and suggestions for ways forward [version 1; peer review: 2 approved with reservations] Wellcome Open Research 2019, 4:172

https://doi.org/10.12688/wellcomeopenres.15425.1

**First published:** 12 Nov 2019, 4:172 https://doi.org/10.12688/wellcomeopenres.15425.1
### List abbreviations

AC - Antenatal Corticosteroid  
CDISC - Clinical Data Interchange Standards Consortium  
CONABIOS - National Council on Bioethics in Health  
EDCTP - European and Developing Countries Clinical Trial Partnership  
IDD - Infectious Diseases Data Observatory  
MESCyT - Ministry of Higher Education, Science and Technology  
MNH - Maternal & Newborn Health  
MoH - Ministry of Health  
NIH - National Institutes of Health  
NICHD - National Institute of Child Health and Human Development  
N-DASH - NICHD Data and Specimen Hub  
REC - Research Ethics Committees  
TDR - Special Programme for Research and Training in Tropical Diseases hosted at the World Health Organization, and sponsored by the United Nations Children’s Fund, the United Nations Development Programme, and the World Bank and WHO.  
WHO - World Health Organisation  
WWARN - WorldWide Antimalarial Resistance Network

### Disclaimer

The views expressed in this article are those of the author(s). Publication in Wellcome Open Research does not imply endorsement by Wellcome.

### Introduction

Data sharing is increasingly mandated by health research funders and publishers\(^4\). Rationales for sharing data include maximising the cost-effectiveness and utility of primary data-sets, minimising duplications and improving the transparency of research with the ultimate aim of progressing science and improving human health\(^1\). However, despite the many mandates, the volume of data shared and reused remain low\(^1\). The reasons for the poor uptake to date need to be better understood\(^1\). Many have cautioned that there are also potential harms in data sharing, such as inadequate consent procedures, breaches of participant confidentiality, group harms of discrimination and stigmatization, and the risk of misinterpretation and wasted resources when data are shared only as a “tick-box exercise” without the necessary accompanying information for data to be accurately interpreted\(^1\). Data sharing and re-use is particularly limited among researchers in low-resource settings\(^1\). This has been attributed to lack of capacity in data science as well as limited funding and lack of protected time for research\(^1\). In addition, many institutions have not yet established policies and infrastructure to undertake data management and facilitating data sharing\(^1\).

In the context of data collected in low-resource settings, concerns have been raised regarding inequitable opportunities to engage in secondary use of data between researchers in well-funded and resource-limited settings\(^1\). What is equitable sharing and how equitable do we need to be? In this context, we present three case studies and their issues related to equity and suggest ways forward. These case studies were discussed at the 2018 Global Forum for Bioethics in Research in Stellenbosch, South Africa focusing on the theme “The ethics of data sharing and biobanking in health research”.

### Case study 1: the antenatal corticosteroids trial

#### Background

Preterm birth has been attributed to 28% of neonatal deaths worldwide\(^1\). Antenatal corticosteroids (ACS) administered to women with high risk of pre-term birth has been known to reduce neonatal mortality in high-resource settings. However, ACS are not routinely used in low-resource settings\(^9\).

The “Antenatal Corticosteroids Trial” was initiated to address this problem\(^9\). The trial was an 18-month two-arm parallel cluster randomized trial designed to assess the feasibility, effectiveness, and safety of an intervention package versus a control group to increase the use of ACS in low- and middle-income countries (LMICs). The study which was funded by the National Institute of Child Health and Human Development (NICHD) under the National Institutes of Health (NIH), USA and was conducted at seven study sites - one site each in Argentina, Zambia, Guatemala, Pakistan, Kenya, and two sites in India.

The details of the trial, methods and results have been described elsewhere\(^9\). Briefly, the intervention arm included health-provider training to identify women who are at risk of preterm birth and provision of a kit to facilitate appropriate use of antenatal corticosteroids. The control arm was standard care in those communities. In both study arms, health providers were trained in the basics of care for low birth weight babies.

To reduce bias, the primary outcome data which was 28-day neonatal mortality among infants less than the 5\(^{th}\) percentile for birthweight, were collected independently by the maternal and newborn health (MNH) registry staff. Secondary outcomes were the level of use of antenatal corticosteroids and suspected maternal infection. The MNH routinely collects data consists of outcome data for all pregnant women residing within the study clusters.

The intervention showed an increase of ACS use to 45% in women delivering infants less than the 5\(^{th}\) percentile for birthweight, compared with about 10% in women in the control group\(^7\). The intervention resulted in an increase in neonatal deaths (3.5 per 1000 livebirths) and an increase in perinatal deaths (5.1 per 1000 births) in the population. In addition, the intervention was also associated with a 3.6% absolute increase in suspected infection among mothers of less-than-5\(^{th}\) - percentile infants and a significant 0.8% increase among all women\(^9,22,23\).

Analyses showed that ACS contributed to the overall increase in neonatal deaths\(^22,23\). One explanation was that the screening approach used to determine risk of preterm birth was not very specific. That could have led to potentially harmful use of ACS for infants not delivered preterm. However, researchers could not make a definitive statement about the impact of the intervention on stillbirth rates in smaller and earlier gestational age fetuses due to the poor gestational age dating available to those participating in the trial\(^22,24\).
In summary, the Antenatal Corticosteroids Trial presented negative results. The intervention employed in the trial did not reduce neonatal mortality in less-than-5th-percentile infants. In addition, the intervention increased deaths in the overall population and increased the risk of maternal infectious deaths.

Selected ethical concerns and suggestions for ways forward
Since Antenatal Corticosteroids Trial presented negative results, a keen interest was generated among the different funding agencies and researchers. The NIH policy expects researchers of primary study who are funded by the NIH to share their individual level de-identified data through the NIH data sharing repositories that make the data accessible for reuse around the world.

Researchers encountered the following ethical issues related to data sharing after the completion of the trial:

1. The Antenatal Corticosteroids Trial was completed in March 2014. According to the NIH data sharing policy, researchers need to share the trial data after the primary publication for further secondary analyses. In this case, it was difficult to abide by this policy since the outcomes of the Antenatal Corticosteroids Trial were captured in the MNH registry which remains an ongoing study. The MNH registry started in 2008 by NICHD Global Network and since then it has been continued as a population-based registry to document maternal and newborn mortality as well as their trends over time. Should data be available for only completed studies or even ongoing studies as well?

After discussion among the primary researchers and the funding agencies, it was decided that the researchers would release the raw data of MNH study (which captures the outcomes of Antenatal Corticosteroids Trial) for the completed period from 2010 to 2013 in NICHD Data and Specimen Hub (NDASH). Since the MNH registry is an ongoing study, the decision of releasing the data of further years will be taken by the primary researchers after periodically conducting the primary trend analysis.

2. Concern was raised on the consent taken which was only for the primary analysis. Since the primary researchers did not plan for sharing of data for secondary analyses during the protocol development, there was no mention in the consent form for secondary analyses of the data. There was no consensus on whether a reconsent was needed during secondary analyses.

3. The primary researchers also had apprehension on the wrong interpretation of the data from the researchers during any secondary analyses. The primary researchers decided that they should be involved in the designing of future secondary analyses to prevent any misinterpretation of the data.

Case study 2: Health research in the the Dominican Republic
Background
The Dominican Republic is a middle-income country with a population of about 10 million. Approximately 74% of the population is covered by a health insurance, the health expenditure is about 6% of the gross domestic product, and the percentage of out-of-pocket health expenditures is about 44%. The Ministry of Higher Education, Science and Technology (MESCyT) is the main public funding agency for biomedical research, however, most of the health research projects are funded by the international pharmaceutical industry. The Ministry of Health (MoH) is responsible for establishing health research policies and priorities. The National Council on Bioethics in Health (CONABIOS, in Spanish) is the authority to approve or reject research protocols. Local research ethics committees have been in place since the early 80s, however, they are not subject to any regulations as health research itself is not regulated by a law but only through an administrative disposition.

In the Dominican Republic, most health research activities are conducted by the international pharmaceutical industry, other international institutions and universities. The implication of this trend is that funds are not allocated towards the diseases and conditions affecting the most vulnerable nor are they directed towards improving outcomes of the healthcare system, and policy development. At the same time, local personnel are contracted as ‘principal investigators’ when in practice they are only dealing with data collection. Where research is conducted by pharmaceutical companies, confidentiality requirements are in place to protect industry rights and the data are not shared with local researchers nor do they participate in data analyses.

This systematic neglect to build research capacity has real consequences. For instance, in 2016 the Dominican Republic reported one of the largest Zika virus outbreaks in the Americas. The first case of Zika was confirmed in January 2016 and decreased by May 2017. Incidence of Guillain-Barré Syndrome was high, however most of the cases had an uncomplicated course. Yet despite the scale of the outbreak in the Dominican Republic, national researchers were not participating as meaningful collaborators.

The National Research Ethics Committees Survey
The National Research Ethics Committees Survey was implemented to identify the number of existing research ethics committees (RECs) in the Dominican Republic, their compositions, organization, activities, ethics review and decision-making processes. Around 400 organizations including health care organizations, academic and research based institutions, both from the public and private sector were contacted. The data collection took place from March 2017 to September 2018. A total of 25 RECs were identified and 20 REC representatives were interviewed using a semi-structured questionnaire with questions about their written policies, composition, activities of REC and ethics review practices such as requesting from researchers a data sharing plan.

The study showed that in the last decade, the number of REC’s increased over 3-fold, from 7 in 2009 to 25 in 2018, half of them from public institutions. Of these, 70% of them have written policies, 30% review clinical trials, 40% meet only twice a year and 45% approved protocols in the first meeting. The study also
found none of the RECs involved mentioned that they request a data sharing plan as part of their ethics review practices. Their written policies did not include requirements for data sharing.

**Selected ethical concerns and suggestions for ways forward**

1. The survey showed that RECs in the Dominican Republic were not requesting data sharing plans as part of their review process. Would it not be reasonable for RECs to request a data sharing plan even though it is not a legal requirement at the moment? Even when many international ethics guidelines suggest that there are compelling reasons to share data, it is still not clear in which instances a REC will have the authority to request a data sharing plan or even mandate data sharing. CONABIOS has the authority to do so, but they do not yet have any policy in this regard.

We think that it would be beneficial for RECs in LMICs to request information regarding data ownership, data management and data sharing as part of their review. In some instances, data sharing should be mandated, for example in research that is looking to solve important local public health issues.

2. In the Dominican Republic, there is a lack of technological and data science capacity to analyse secondary data. Sharing of data with local researchers who do not have the capacity to analyse the data will not be beneficial. In this regard, we think that capacity building (and retention) is necessary in order for LMIC researchers to benefit from the research and the data collected. For example, a local data scientist or statistician could be included as part of the research team. International collaborative work should include the local research teams in all phases of the research project, not just in the data collection phase.

CONABIOS should offer guidance to REC in terms of policies and standards on data sharing. The MoH and the MESCyT should work together in the development of a platform for data access including consideration on policy development, organizational structure, central platforms (local and regional) to access and analysis the data, and capacity building agreements.

**Case study 3: the worldwide antimalarial resistance network (WWARN)**

**Background**

WWARN was established in 2009 to understand and curtail the threat of antimalarial resistance. Key to the delivery of WWARN’s aims was engaging with global malaria researchers and encouraging them to share their data with the central WWARN repository, at a time before data sharing was required by any funders, publishers or regulatory agencies. The real and perceived challenges to data sharing were many and diverse, so WWARN developed a number of strategies to enable and encourage equitable sharing of robust data to inform malaria treatment policies and practices. This case study will focus on efforts to promote equity in sharing of data by, and with, researchers from malaria-endemic countries.

Malaria is a poverty-related disease, and its control and eventual eradication are threatened by the spread of parasite resistance to all currently available antimalarials, including the pivotal artemisinins that have played a central role in global decreases in malaria burden since 2000. Promptly sharing reliable data on the efficacy of antimalarial medicines has the potential to prevent or slow antimalarial drug resistance. However, requests to share data to address this critical global health threat have resulted in expressions of concern from researchers, including that the quality of data may be scrutinised or study outputs challenged by external researchers, and that researchers in low-resource malaria-endemic settings are less able to benefit from the fruits of data sharing than researchers in better resourced settings.

Over the past decade WWARN has worked with collaborators in over 280 institutions globally to develop and update its scientific, technical, ethical and governance frameworks to promote equity in data sharing. Key aspects of these efforts which address the primary concerns of the malaria research community are capacity strengthening and technical support in data standardisation and quality, as well as inclusion of primary data generators in secondary analyses.

The impact of these efforts is demonstrated by the size of the WWARN platform which, thanks to the contributions of the global malaria research community, now holds over 80% of the world’s individual patient clinical trial data on artemisinin-based combination antimalarials. These data on factors affecting the efficacy of antimalarial medicines have been used to optimise treatment regimens for high-risk groups including pregnant women, young and malnourished children, and provides evidence to inform the development of new antimalarial drugs.

**Selected ethical issues and suggestions for ways forward**

1. In order to address the concerns of many researchers, and not just those based low-resource settings, that their raw data may not be entirely ready for international scrutiny and their study outputs challenged, WWARN has invested heavily in providing researchers with the resources needed to feel more confident in the quality of data that they share.

a) WWARN developed and continues to expand its tools and resources to enhance the efficiency and quality of planning, executing, analysing and reporting primary data collection (see WWARN Tools and Resources page).

b) This is supported by WWARN’s external quality assurance and proficiency testing programme, to enhance data quality and comparability for laboratories conducting antimalarial drug assays.

c) The WWARN Informatics platform accepts data submitted in almost any format, with the related protocol / case report forms / metadata / data dictionaries needed to ensure that data are usable for secondary analyses. The contributed data are curated and standardised using established data and statistical management plans. The “data contributor” receives a study report which includes a list of changes made during curation and processing.
and a list of any outliers or unexpected results. The original data files and the resultant data set that complies with the Clinical Data Interchange Standards Consortium (CDISC) standards (where applicable) are stored in the WWARN repository, which is an re3data registered repository. These outputs are all available to the contributor or designee, enhancing the quality of their datasets for their own future use.

2. In order to address the concerns of many researchers, primarily in low-resource settings, that they may be less able to benefit from the fruits of data sharing than researchers in better resourced settings, WWARN has developed a number of strategies.

a) In order to give data contributors more choice about how their data can be accessed, WWARN has recently changed its governance frameworks. Data contributors can now choose between “contributor controlled access” where the contributor will review each individual request, or for this to be done through the WHO hosted independent Data Access Committee.

b) WWARN also organizes study groups to bring together data contributors conducting individual participant data meta-analyses to answer important research questions that cannot be answered as reliably or efficiently by individual studies or aggregate data meta-analyses.

Examples of impactful meta-analyses that informed improvements in the treatment of uncomplicated malaria with the following artemisinin-based combination treatments:

- Dihydroartemisinin-piperaquine: Improved dosing recommendations in young children;\(^\text{2,3}\)  
- Artemisinin-amodiaquine: The enhanced efficacy of the fixed dose combination relative to loose tablets;\(^\text{11}\) and  
- Artemether-lumefantrine: Sub-optimal lumefantrine exposure in malnourished children\(^\text{15}\).

A research question can be proposed by anyone, and researchers from malaria endemic countries may be best placed to identify important knowledge gaps. These study groups not only benefit from pooling the individual patient data shared, but also from skill-sharing of the expertise of each of the primary researchers and technical and statistical support provided through the WWARN data platform. Depending on each study group member’s level of engagement in the secondary analysis, the members are authors, collaborators, or acknowledged in resulting publications\(^\text{46}\).

3) Increasing capacity building efforts to enable researchers from malaria-endemic LMICs to be able to access and use secondary data to answer questions of importance to malaria and other NTD control and elimination efforts. These include online open access resources, training workshops conducted in East, West and Southern Africa, and to date hosting ten EDCTP/TDR career development fellows from LMICs to gain the skills required to lead future efforts to make the best use of available data to inform policy and practice. As a part of the Infectious Diseases Data Observatory (IDDO), WWARN also contributes to work with other research communities to replicate this model for other neglected, poverty-related diseases and emerging infections.

**Discussion and recommendations**

The recent requirements for data sharing by an increasing number of funders, publishers and regulatory agencies risk exacerbating existing inequities between researchers in high-resource and low-resource settings, and data reuse is unlikely to produce the expected public health benefits unless critical challenges are addressed. The case studies in this paper provide concrete examples of real challenges and some potential solutions related to equitable data sharing. We recommend the following ways forward:

1. **Capacity building**

Planning and collecting good quality data requires significant investment in terms of expertise, experience, skills, time and effort on the part of primary data collectors. In light of this, specific funding should be allocated for capacity building programmes to improve data management as well as data reuse capacity in researchers in low-resource settings. Collaborations between researchers in high- and low-resource settings as a condition for sharing may strengthen such capacity building efforts. Collaborations with the primary researchers are especially important where interpretation of the data requires in-depth understanding of the population the data are drawn from and the context in which the data were collected and curated\(^\text{10}\). Initiatives including those led by WWARN (Case study 3) demonstrate that equitable sharing can be achieved, following considerable investment in human resources, technology and infrastructure for the curation and sustainable sharing of research outputs. Efforts to develop data management and data sharing courses that will be made freely available are underway.

2. **Investments**

Funding in data management and sharing platforms supporting poverty-related disease research communities should be increased. Designated funding should be included in research grants of the primary study that budget for costs and time spent on an activities specific to data sharing, such as the additional curation needed, data storage, staff time, hardware and software. Investments are also needed in the management of platforms supporting complex data integration and analyses. Without these investments, the recent requirements for data sharing by an increasing number of funders, publishers and regulatory agencies risk exacerbating inequities between researchers in well-resourced and resource-limited settings, and data reuse is unlikely to produce the expected public health benefits.

3. **Data sharing policies**

Although data sharing has been widely promoted and researchers have increased their data sharing activities, very few research groups and institutions have formal data sharing policies\(^\text{11}\). Institutional data sharing policies are important for many reasons:
for members of the institution to have a shared understanding of their own data sharing processes, to safeguard the interests of their researchers as well as those of their data subjects. The data sharing policy should provide guidelines for secondary users to request for data and what are the priority secondary analyses such as those that are consistent with institutional aims. It should also include when special conditions of access should be put in place such as requirements for collaborations on secondary analyses. In addition, an institution may set embargo periods, preferential access provisions (e.g. to collaborators and LMIC researchers, and to secondary analyses that directly benefit communities that generated the primary data). These policies should take into account their context, type of data and database and relevant existing regulations and policies (e.g. funders)\(^1\). For example, in the case of the Maternal & Newborn Health Registry (Case Study 1), the policy may state that data underlying the study published will be shared, and not the entire registry.

4. Incentives and attributions
In order to avoid disincentivising primary research, appropriate recognition and credit should be provided to primary researchers and their teams\(^9\). In light of current developments in data sharing, mainstream international guidelines on authorship criteria should be revisited. The current International Committee of Medical Journal Editors may not be adequate to account for the different levels and types of contributions of the primary researchers in secondary analyses. The discussions and decisions around authorship should involve both primary and secondary researchers including those in low-resource settings\(^1\). Creative solutions have been suggested such as the “CRediT taxonomy” system and “data authorship” but these have not been widely accepted\(^9,10\). While the CRediT taxonomy system specify roles of authors, it does not provide guidance on when an individual qualifies to be an author. Data authorship is not yet held in the same academic kudos as manuscript authorship.

5. Consent and community engagement
For new studies, researchers should ensure that participants have given consent for sharing their data with researchers external to the primary research team. ‘Broad consent’ for unspecified future use is currently the most widely accepted mechanism to obtain participant consent for sharing data beyond the primary research teams\(^5,10\). Research staff who are tasked to obtain broad consent must be appropriately trained. For multicentre studies, it is necessary to engage with collaborators to ensure that clinical study agreements include provisions for data sharing and obtaining appropriate consent.

For primary studies, what is appropriate information and what constitutes adequate understanding on the part of potential research participants remain enduring ethical questions\(^41,42\). Studies have shown that communications about data sharing adds another layer of complexity to the informed consent process\(^10\). Community and public engagement may help to improve general understanding of data sharing among research communities. Such engagement is also important to discern what constitutes sensitive data, what secondary uses might cause harm or stigma to communities, and what limitations should be placed on sharing with external parties. A combination between conventional engagement approaches such as holding public talks and consultation with community advisory boards\(^31,44\), and creative initiatives such as arts-science collaborations and café-style talks\(^45–49\), may be necessary to refine both the development of core information about data sharing to be provided to all research participants, and appropriate solutions for context specific-challenges arising when explaining data sharing.

Conclusions
To promote equitable data sharing, the interests of multiple research stakeholders must be considered and including: (1) the primary researchers, their wider teams, and their institutions, (2) the primary study participants and their communities, (3) secondary users, their wider teams, and their institutions and (4) the broader public that stand to benefit from the knowledge generated through research studies. Equitable data sharing requires investments and efforts from all stakeholders involved. We need to go beyond merely minimising harms to research participants and increase the promotion of the interests of their communities by encouraging data sharing and re-use while protecting the interests of primary researchers and their institutions.

Data availability
Underlying data
No data are associated with this article.

Acknowledgements
We thank Jantina de Vries, Athula Sumathipala and Adrienne Hunt for their helpful suggestions on the case studies and input in the manuscript.

References
1. Taichman DB, Sahni P, Pinborg A, et al.: Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. Lancet. 2017, 389(10086): e12–e4. Published Abstract | Publisher Full Text
2. Wellcome Trust: Policy on data, software and materials management and sharing. 2017. Accessed 14 April 2019. Reference Source
3. European Medicines Agency: Policy on publication of clinical data for medicinal products for human use (EMA/240810/2013) 2014. Accessed 14 April 2019. Reference Source
4. US National Institutes of Health: Final NIH Statement on Sharing Research Data 2003. Reference Source
5. Bull S, Cheah PY, Denny S, et al.: Best Practices for Ethical Sharing of
Open Peer Review

Current Peer Review Status: ?  ?

Version 1

Reviewer Report 24 February 2020

https://doi.org/10.21956/wellcomeopenres.16869.r37714

© 2020 Nicol D. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Dianne Nicol
Centre for Law and Genetics, University of Tasmania, Hobart, Australia

I support publication of this manuscript, although I recommend that some amendments should be undertaken prior to indexing.

I found the introduction was too short and glossed over some of the complexities of data sharing. I think that some points need further discussion. For example, could the authors please address the following questions? If data sharing is mandated, why does the volume remain low? Is it too soon to be thinking about data sharing, given the problems associated with lack of capacity and lack of infrastructure? Why were these three case studies selected?

I found that the case studies (particularly case study 1) tended to focus a little too much on the studies themselves and not enough on the topic of the manuscript – data sharing. It seemed to me that at least some of the issues with low-resource settings relate to the subject matter of the case studies, rather than the issues to do with data sharing per se. I found this to be particularly the case for the first case study, where the primary issue was the intervention itself. Furthermore, the three points raised under ethical concerns associated with this study did not seem to me to be specific to low-resource settings. Rather, then raised more general concerns about sharing of clinical data. Is the point simply that the trials were conducted in low-resource settings? If they were conducted in higher-resource settings, would the outcomes have been different?

The second case study raised issues more squarely focused on data sharing. The primary issue here was the inadequacy of the ethics review system, particularly with regard to data sharing plans.

The third case study is also squarely focused on data sharing, in this instance sharing of data on antimicrobial resistance. I found that some of the points raised in the ethical issues associated with this case study to be highly specific and technical. Wouldn’t it be better to reduce the number of points relating to ethical issues in this case study and make them about data sharing more generally? Reporting on the specific recommendations associated with WWARN could then be undertaken elsewhere.
I have no particular concerns with the discussion and strongly support the five recommendations.

With regard to style and structure of the manuscript, I found that the three case studies read as if they had been written by three different authors (perhaps because they were). It would be useful for the authors to spend a little more time ensuring that there is greater harmonisation in style. Some of the sentences required reading several times. One example is the sentence starting: ‘The intervention showed an increase of ACS ...’. I still don’t quite understand what it is trying to say.

Case studies 2 and 3 could do with a little more work on sentence construction throughout. For example, this sentence didn't make sense to me: ‘The first case of Zika was confirmed in January 2016 and decreased by May 2017’. What was it that decreased?

Some of the sentences are much too long. See, for example: ‘In order to address the concerns of many researchers, and not just those based low-resource settings, that their raw data may not be entirely ready for international scrutiny and their study outputs challenged, WWARN has invested heavily in providing researchers with the resources needed to feel more confident in the quality of data that they share.’

There also tends to be overuse of acronyms – are they all necessary? Careful editing of the text by the authors would make the manuscript more readable and would encourage greater engagement by the target readership.

Is the rationale for the Open Letter provided in sufficient detail? Partly

Does the article adequately reference differing views and opinions? Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations? Yes

Is the Open Letter written in accessible language? Partly

Where applicable, are recommendations and next steps explained clearly for others to follow? Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Ethical, legal and social implications of genomics and related technologies.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
“Equitable data sharing: challenges and suggestions for ways forward” is an open letter presenting three cases and recommendations. While cases are informative, I have doubts regarding the nature and the structure of the manuscript. The manuscript has an unusual structure – three cases followed by general recommendations. I am not sure if this should be submitted as an “Open Letter” or rather a series of cases or developed and published as a full-length manuscript or guideline.

Another concern is that a general introduction does not prepare the reader to very specific and detailed cases. I also don’t feel that the final recommendations – while clear and reasonable – are supported by just three cases. Finally, my impression is that cases present too many distractive details, not necessarily important and connected with the main topic. For example a very short case of The Antenatal Corticosteroids Trial mentions four different dates: 2014, 2008, 2010, 2013. I am not sure if this is really important information here. Moreover the letter has 15 acronyms. Some are used no more than 2-3 times. I found this unnecessary and very distractive.

Minor:
“Health research in the the Dominican Republic” – delete double “the”.

Is the rationale for the Open Letter provided in sufficient detail?
Partly

Does the article adequately reference differing views and opinions?
Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language?
Partly

Where applicable, are recommendations and next steps explained clearly for others to follow?
Yes
**Competing Interests:** Dr. Waligora reports personal fees from the Advisory Bioethics Council, Sanofi outside the submitted peer-review report.

**Reviewer Expertise:** Research ethics, bioethics.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.