

## Perspective Piece

### Strength in Numbers: The WWARN Case Study of Purpose-Driven Data Sharing

Georgina S. Humphreys,<sup>1,2</sup> Halidou Tinto,<sup>3</sup> and Karen I. Barnes<sup>4,5\*</sup>

<sup>1</sup>WorldWide Antimalarial Resistance Network (WWARN), University of Oxford, Oxford, United Kingdom; <sup>2</sup>Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom; <sup>3</sup>Institut de Recherche en Sciences de la Santé—Clinical Research Unit of Nanoro (IRSS-CRUN), Ouagadougou, Burkina Faso; <sup>4</sup>Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa; <sup>5</sup>WorldWide Antimalarial Resistance Network (WWARN) Pharmacology and Southern African Regional Hub, University of Cape Town, Office K27-24 Old Main Building Groote Schuur Hospital, Anzio Road, Observatory, 7925, Cape Town, South Africa

**Abstract.** Data are the basis for all scientific output. The sharing of data supporting that output is an important aspect of scientific communication, and is increasingly required by funders and publishers. Yet, academic advancement seldom recognizes or rewards data sharing. This article argues that although mandating data sharing will increase the amount of data available, this will not necessarily enable or encourage the secondary analyses needed to achieve its purported public good. We, therefore, need to build models that maximize the efficiency of processes for data collation and curation, and genuinely reward those engaged in data sharing and reuse. The WorldWide Antimalarial Resistance Network has 10 years of experience as a data platform, and its study group approach provides an example of how some of the challenges in equitable and impactful data-sharing and secondary use can be addressed, with a focus on the priorities of researchers in resource-limited settings.

Data sharing is important, and increasingly mandated by funders and publishers, which will inevitably result in an increased number of datasets being shared. However, more now needs to be done to ensure that sharing these data results in the expected benefits of increased scientific productivity, efficiency, and public health impact. In the context of data collected in low- and middle-income countries, concerns have been raised regarding inequitable opportunities to engage in secondary use of data between researchers in well-resourced and resource-limited settings. This is of particular importance for research in poverty-related diseases, such as malaria, neglected tropical diseases, and emerging infections such as Ebola. In such fields, differences in technical and financial capacity between primary data producers and secondary data users are often large. Although primary data originate in resource-limited settings, it is in high-income settings where resources are more readily available for conducting secondary analyses.<sup>1,2</sup>

Meta-analysis based on aggregated information extracted from publications remains the most widely used method to synthesize evidence to guide policy recommendations. However, results based on published aggregated data alone may be limited and potentially misleading because of publication bias and heterogeneity in methodologies.<sup>3</sup> Data are inconsistently reported, with key information often missing.<sup>4</sup> Heterogeneity of methodologies often restricts the feasibility of the meta-analysis approach.<sup>5</sup> For example, an aggregated data meta-analysis on early parasitological response following treatment with artemisinin-containing regimens found that the proportion of patients remaining parasitemic at 24, 48, and 72 hours was reported in less than half of the relevant publications, highlighting the need for individual patient data (IPD) meta-analyses.<sup>6</sup>

Individual patient data meta-analysis is an increasingly recognized alternative method to evidence synthesis that enables standardization of outcomes across studies, and may improve control of bias.<sup>3,7,8</sup> Unlike most aggregated meta-analyses, the IPD approach is often applied beyond randomized controlled trials (RCTs). This allows for inclusion of other study designs that may be more representative of the normal context of use. Individual patient data also gives greater statistical power and analytical flexibility—for example, the ability to study drug efficacy for a particular subgroup of patients who are an important target population, but usually underrepresented in individual clinical trials. Huang et al.<sup>9</sup> compared results from aggregated meta-analyses and IPD meta-analyses and found that among the 204 paired studies, 91.7% agreed in the overall effect, but the IPD meta-analyses enabled seven-times more subgroup analyses and identified 14 times more potential interactions.

Notwithstanding the advantages of the IPD approach, using IPD from clinical trials can raise ethical, logistical, and methodological challenges. These may include reuse for purposes not originally intended when consent was agreed, inadequate recognition of the primary generators of the data, and regulatory issues such as national policies that limit data sharing. The challenges of collating, curating, and analyzing IPD are significant, and certainly more complex and resource-consuming than for aggregate data. This is reflected in the relative scarcity of IPD meta-analyses in the literature; in a survey of corresponding authors of meta-analyses of RCTs in general medicine, only 4.2% included IPD.<sup>10</sup>

In other health disciplines, such as genomics, the culture of data sharing and reuse is better established, with clear evidence of a citation benefit for studies with a publicly available dataset.<sup>11</sup> However, similar evidence from clinical and epidemiological studies is limited. Although there are some recent examples for controlled access data sharing and analysis of clinical trials data, these largely focus on data collected in resource-rich settings where data management standards make it easier to share, curate, and combine datasets.<sup>12,13</sup>

Individual patient data meta-analyses are particularly important in the field of poverty-related infectious diseases that

\*Address correspondence to Karen I. Barnes, Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa and WorldWide Antimalarial Resistance Network (WWARN) Pharmacology and Southern African Regional Hub, University of Cape Town, Office K27-24 Old Main Building Groote Schuur Hospital, Anzio Road, Observatory, 7925, Cape Town, South Africa. E-mail: karen.barnes@uct.ac.za

have relatively low levels of research funding, and so relatively few data are available. Thus, the potential for new findings from bringing these sparse data together is much higher. Reviews of the published literature on malaria, schistosomiasis, soil-transmitted helminths, and visceral leishmaniasis highlighted the importance of IPD meta-analyses to not only increase statistical power, but also to study specific sub-populations whose outcomes are often hidden in aggregated data,<sup>14,15</sup> thus ensuring that the full public health benefit of these studies is realized.

There are a number of ways to ensure or encourage sharing: to pay for access to data, to offer the incentive of co-authorship of the secondary use publication, and, most often, funders and publishers mandating data sharing. However, the current academic culture often runs counter to data sharing, as it promotes individual achievement and encourages academics to be proprietorial. Enforcing data sharing, without incentives, may merely compel researchers to meet only a minimum set of requirements, and not fully engage in sharing all information needed to make meaningful secondary use possible.<sup>16</sup>

More creative models of data sharing and secondary analysis are needed. Although there has been much discussion about potential benefits and concerns around data sharing, particularly for those researchers based in low- and middle-income countries,<sup>1,16</sup> there are very few examples of successful models for purpose-driven data sharing and secondary analysis. To address this gap, the WorldWide Antimalarial Resistance Network (WWARN) launched its first study group in 2011, which brought data contributors together to conduct purpose-driven, equitable IPD meta-analyses with support from core WWARN staff. The study groups were driven by a shared interest in making the best use of available data for answering a specific question, and the commitment to encourage full involvement and ensure recognition for data contributors. Study group membership is open to anyone with relevant data. This approach was designed not only to ensure appropriate recognition of those researchers who invest significant resources in collecting the data, but also to benefit from the participants' scientific expertise and detailed understanding of their data and the context in which it was collected. Depending on each study group member's level of engagement, the members are authors, collaborators, or personally acknowledged in resulting publications. Much has been learnt through WWARN's pioneering efforts; this has been independently evaluated in a case study by Pisani and Botchway.<sup>17</sup>

WorldWide Antimalarial Resistance Network increases efficiency and reduces data loss by standardizing and storing datasets for multiple reuse without needing repeated data curation. By combining IPD from studies conducted by many different groups and in many different countries and settings, it has been possible to observe patterns in malaria treatment efficacy that would not have been discernible in an individual study. The nine publications generated by study groups<sup>18–26</sup> up to December 2017 have included a total of 244 authors, with contributions from 189 separate study datasets shared through the WWARN platform. More than 60% (114) of the studies used in these nine pooled analyses have been included in more than one study group, with 36 (19%) study datasets used in four or more analyses.

An example of the impact of WWARN's IPD meta-analyses is a report that included more than 7,000 patients given

dihydroartemisinin–piperaquine for the treatment of uncomplicated malaria from 26 studies.<sup>21</sup> The pooling of these data showed generally high cure rates, with only 136 treatment failures, but that young children were three times more likely to fail treatment than older patients because they were suboptimally dosed. These findings were subsequently supported by pharmacometric models,<sup>27</sup> and ultimately contributed to dosage recommendation changes in the WHO treatment guidelines.<sup>28</sup> Another example is the Parasite Clearance study group<sup>23</sup> that pooled IPD from more than 6,900 patients enrolled in 24 studies with frequent parasite counts. This pooled analysis was used to validate a tool developed to standardize estimates of the parasite clearance rates that has been cited by more than 100 subsequent peer-reviewed publications.<sup>29</sup> WorldWide Antimalarial Resistance Network is building on this experience to understand how best to encourage more malaria endemic country-based researchers to lead study groups, including providing more tools, infrastructure, and technical support to enable them to answer the questions most relevant to them. We should continue to innovate by experimenting with different data sharing models, particularly as data sharing practices develop and mechanisms for incentivizing and recognizing data sharing are established.

To realize the full potential of shared data, it is imperative that funders work together with researchers to launch funding calls specifically targeted at secondary analysis. In addition, long-term infrastructure support is required to efficiently curate and store reusable data. Given the challenges of data collection and collation, we have a responsibility to retain its value and maximize its impact. If we can create trusted environments, such as WWARN, where skills and knowledge are shared, those contributing are recognized and inefficiencies are reduced, then the benefits of data sharing will be enhanced and more evenly distributed. By embracing a purpose-driven equitable approach to data sharing, the potential to learn more from the data already generated by staff and patients around the world will be unlocked.

Received August 8, 2018. Accepted for publication September 16, 2018.

Published online October 22, 2018.

Authors' addresses: Georgina S. Humphreys, WorldWide Antimalarial Resistance Network (WWARN), University of Oxford, Oxford, United Kingdom, E-mail: georginahumphreys@gmail.com. Halidou Tinto, Institut de Recherche en Sciences de la Santé-Clinical Research Unit of Nanoro (IRSS-CRUN), Ouagadougou, Burkina Faso, E-mail: halidoutinto@gmail.com. Karen I. Barnes, Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa, and WorldWide Antimalarial Resistance Network (WWARN) Pharmacology and Southern African Regional Hub, University of Cape Town, Cape Town, South Africa, E-mail: karen.barnes@uct.ac.za.

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 author and source are credited.

## REFERENCES

1. Bull S, Roberts N, Parker M, 2015. Views of ethical best practices in sharing individual-level data from medical and public health research: a systematic scoping review. *J Empir Res Hum Res Ethics* 10: 225–238.
2. Serwadda D, Ndebele P, Grabowski MK, Bajunirwe F, Wanyenze RK, 2018. Open data sharing and the global south—who benefits? *Science* 359: 642–643.

3. Stewart LA, Parmar MK, 1993. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 341: 418–422.
4. Mills E, Wu P, Gagnier J, Heels-Ansdell D, Montori VM, 2005. An analysis of general medical and specialist journals that endorse CONSORT found that reporting was not enforced consistently. *J Clin Epidemiol* 58: 662–667.
5. Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, Burchill SA, 2003. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *Br J Cancer* 88: 1191–1198.
6. Das D, Price RN, Bethell D, Guerin PJ, Stepniewska K, 2013. Early parasitological response following artemisinin-containing regimens: a critical review of the literature. *Malar J* 12: 125.
7. Chalmers I, 1993. The cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci* 703: 156–165.
8. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, Rovers M, 2015. Individual participant data (IPD) meta-analyses of randomised controlled trials: guidance on their use. *PLoS Med* 12: e1001855.
9. Huang Y, Tang J, Wai-San Tam W, Mao C, Yuan J, Di M, Yang Z, 2016. Comparing the overall result and interaction in aggregate data meta-analysis and individual patient data meta-analysis. *Medicine (Baltimore)* 95: e3312.
10. Kovalchik SA, 2012. Survey finds that most meta-analysts do not attempt to collect individual patient data. *J Clin Epidemiol* 65: 1296–1299.
11. Piwowar HA, Vision TJ, 2013. Data reuse and the open data citation advantage. *PeerJ* 1: e175.
12. Vivii, 2018. Available at: <http://vivii.org/>. Accessed July 1, 2018.
13. Project Data Sphere, 2014. Available at: <https://projectdatasphere.org/projectdatasphere/html/home>. Accessed July 1, 2018.
14. Julé AM et al., 2016. The schistosomiasis clinical trials landscape: a systematic review of antischistosomal treatment efficacy studies and a case for sharing individual participant-level data (IPD). *PLoS Negl Trop Dis* 10: e0004784.
15. Bush JT, Wasunna M, Alves F, Alvar J, Oliaro PL, Otieno M, Sibley CH, Strub Wourgaft N, Guerin PJ, 2017. Systematic review of clinical trials assessing the therapeutic efficacy of visceral leishmaniasis treatments: a first step to assess the feasibility of establishing an individual patient data sharing platform. *PLoS Negl Trop Dis* 11: e0005781.
16. Merson L, Gaye O, Guerin PJ, 2016. Avoiding data dumpsters—toward equitable and useful data sharing. *N Engl J Med*. 374: 2414–2415.
17. Pisani E, Botchway S, 2016. *Learning from the Pioneers: Lessons about Data Platforms Drawn from the WWARN Experience*. Wellcome Trust. Paper. Available at: <https://doi.org/10.6084/m9.figshare.4476308.v1>. Accessed July 1, 2018.
18. Hoglund RM et al., 2017. Population pharmacokinetic properties of piperazine in falciparum malaria: an individual participant data meta-analysis. *PLoS Med* 14: e1002212.
19. WWARN Artemisinin based Combination Therapy (ACT), Africa Baseline Study Group; Dahal P, d'Alessandro U, Dorsey G, Guerin PJ, Nsanzabana C, Price RN, Sibley CH, Stepniewska K, Talisuna AO, 2015. Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data. *BMC Med* 13: 212.
20. Worldwide Antimalarial Resistance Network (WWARN) AL Dose Impact Study Group, 2015. The effect of dose on the antimalarial efficacy of artemether–lumefantrine: a systematic review and pooled analysis of individual patient data. *Lancet Infect Dis* 15: 692–702.
21. WorldWide Antimalarial Resistance Network (WWARN) AS-AQ Study Group; Adjuik MA et al., 2015. The effect of dosing strategies on the therapeutic efficacy of artesunate–amodiaquine for uncomplicated malaria: a meta-analysis of individual patient data. *BMC Med* 13: 66.
22. WorldWide Antimalarial Resistance Network (WWARN) DP Study Group, 2013. The effect of dosing regimens on the antimalarial efficacy of dihydroartemisinin–piperazine: a pooled analysis of individual patient data. *PLoS Med* 10: e1001564.
23. WWARN Parasite Clearance Study Group; Abdulla S et al., 2015. Baseline data of parasite clearance in patients with falciparum malaria treated with an artemisinin derivative: an individual patient data meta-analysis. *Malar J* 14: 359.
24. Venkatesan M et al., 2014. Polymorphisms in *Plasmodium falciparum* chloroquine resistance transporter and multidrug resistance 1 genes: parasite risk factors that affect treatment outcomes for *P. falciparum* malaria after artemether–lumefantrine and artesunate–amodiaquine. *Am J Trop Med Hyg* 91: 833–843.
25. WorldWide Antimalarial Resistance Network (WWARN) Lumefantrine PK/PD Study Group, 2015. Artemether–lumefantrine treatment of uncomplicated *Plasmodium falciparum* malaria: a systematic review and meta-analysis of day 7 lumefantrine concentrations and therapeutic response using individual patient data. *BMC Med* 13: 227.
26. WWARN Gametocyte Study Group, 2016. Gametocyte carriage in uncomplicated *Plasmodium falciparum* malaria following treatment with artemisinin combination therapy: a systematic review and meta-analysis of individual patient data. *BMC Med* 14: 79.
27. Bergstrand M, Nosten F, Lwin KM, Karlsson MO, White NJ, Tarning J, 2014. Characterization of an in vivo concentration–effect relationship for piperazine in malaria chemoprevention. *Sci Transl Med* 6: 260ra147.
28. WHO, 2015. *Guidelines for the Treatment of Malaria*. Available at: <http://www.who.int/malaria/publications/atoz/9789241549127/en/>. Accessed July 1, 2018.
29. Flegg JA, Guerin PJ, White NJ, Stepniewska K, 2011. Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. *Malar J* 10: 339.