BJOG An International Journal of Obstetrics and Gynaecology

DOI: 10.1111/1471-0528.15393
www.bjog.org

Commentary

COLLECT, a collaborative database for pregnancy and placental research studies worldwide

JE Myers,[a] L Myatt,[b] JM Roberts,[c] CWG Redman[d] For the Global Pregnancy Collaboration (CoLab)

[a] Maternal and Fetal Health Research Centre, Manchester Academic Health Science Centre, Manchester, UK [b] Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, USA [c] Department of Obstetrics and Gynecology and Reproductive Sciences, Epidemiology and Clinical and Translational Research, Magee Women's Research Institute, University of Pittsburgh, Pittsburgh, PA, USA [d] Department of Obstetrics and Gynecology, University of Oxford, Oxford, UK
Correspondence: Dr JE Myers, Maternal and Fetal Health Research Centre, Manchester Academic Health Science Centre, St Mary's Hospital, 5th floor Research, Manchester M13 9WL, UK. Email Jenny.myers@manchester.ac.uk

Accepted 23 June 2018. Published Online 23 August 2018.

Please cite this paper as: Myers JE, Myatt L, Roberts JM, Redman CWG, For the Global Pregnancy Collaboration (CoLab). COLLECT, a collaborative database for pregnancy and placental research studies worldwide. BJOG 2018; https://doi.org/10.1111/1471-0528.15393.

Pregnancy-related complications, pre-eclampsia, fetal growth restriction and preterm birth persist as leading causes of maternal and perinatal death and long-term morbidity. Although the impact of these conditions is significant in high-income countries, the burden of these diseases falls disproportionately on low- and middle-income countries.

Our understanding of the underlying pathophysiology of the obstetric syndromes pre-eclampsia, fetal growth restriction and preterm birth has progressed, but the development of new treatments and preventive strategies has slowed, with very few new interventions successfully translating into clinical practice. There are several reasons why research efforts have stalled, these include, but are not restricted to, the following observations. (1) Individual studies are often small in size with small effect sizes and therefore the findings are less likely to be reproducible.[1,2] (2) even larger studies usually lack sufficient power to study 'pure phenotypes' among the obstetric syndromes and rely on arbitrarily defined clinical observations (blood pressure, birthweight and gestational age), which do not accurately reflect the underlying pathophysiology. (3) The majority of research data has been collected in high-income countries where the severest pregnancy outcomes are less common and often censored by access to antenatal care and obstetric intervention. (4) To date, individual research groups and consortia have collected data using their own specific methodology and data capture tools and often reported different outcome metrics, making comparisons between studies difficult.[3]

It seems obvious that collaborative efforts and data sharing would help to address some of these barriers and facilitate larger, cohesive epidemiological and translational research. In practice, however, there are many practical and bureaucratic barriers to merging data across multiple studies and from different countries. Not all of these are easily overcome, but one practical obstacle relates to the enormous task of standardising variables, particularly outcome variables, across many data sets. This has been illustrated by recent collaborative efforts to merge data sets that have quantified angiogenic factors during pregnancy as predictive or diagnostic markers of placental disease[4] and an ongoing UK-funded study to build risk factor prediction models for pre-eclampsia across multiple data sets.[5] Despite the collaboration between a large number of international investigators and access to data from > 100 000 pregnancies, a disproportionate amount of resources is necessary to clean and combine data before any analysis can begin. In addition, misalignment between variable structures within data sets often results in significant data loss.

The Global Pregnancy Collaboration (CoLab) was established with support from the Bill & Melinda Gates Foundation as an international consortium of investigators and centres to enable sharing of data and biological samples worldwide with the objective of facilitating collaborative research studies of adverse pregnancy outcomes and ultimately improving the health of women and their infants. The team has outlined the standards necessary for data collection, including minimal and optimal data sets for pre-eclampsia,[6] for placental tissue collection[7] and for harmonisation of study design in relation to pre-eclampsia and related research.[3] An important goal of such efforts is to facilitate collaborative research between investigators

from diverse centres, countries and populations. Several other international initiatives have also recently promoted the use of core outcome sets for a wide range of conditions, including pregnancy complications.[8] Standardised reporting of outcomes to enable comparison across studies and meaningful meta-analysis is crucial to the progression of research within our field. Collaboration and discussion between interdisciplinary researchers and the public through advocacy organisations and other networks is essential to the creation of meaningful outcomes. Core outcome sets need to contain measurable variables that capture the impact of a complication or intervention and enable fair comparison of outcomes across studies. A recent example of such an endeavour is the iHOPE project, which has used the Delphi consensus technique to develop a core outcome set for pre-eclampsia.[9]

Having established standardised core outcome variables, the next challenge is how these are to be reliably recorded across studies. The investigators within CoLab have addressed this with the COLLECT database, which is based on a template (MedSciNet) derived by Professor Robyn North and the international SCOPE consortium.[10,11] A suggested list of standardised variables was established through panel discussion by a group of international investigators[6] and variables were retained/deleted from the SCOPE template as appropriate. The design of a generic database is, in itself, associated with significant challenges and it would be naive to expect that a single platform could provide a solution for all pregnancy research worldwide. The COLLECT database, however, can accommodate different study designs and goals by using a modular structure. This format includes a series of standard pages including participant demographics fields, pregnancy characteristics, antenatal visit information, pregnancy complications and birth outcomes. Additional modules (pages) can be added depending on a specific study's needs and this has already been successfully achieved with ongoing studies (ISRCTN 12695929). The database can also be translated and this has already been successfully achieved in the ongoing PREPARE study in Brazil. There is also the possibility of converting existing databases to the COLLECT format for a modest cost (approximately US$ 5000).

A second and very important motivation underpinning the design and creation of the COLLECT database was the aspiration to provide researchers across all resource settings with a robust and secure platform through which they could collect useful and meaningful data in an efficient and cost-effective manner. Developing a study-specific database is expensive and time-consuming and frequently takes resources away from other aspects of research. Maintenance and updates to databases, in addition to their storage, are also costly. The COLLECT database has a web-based data entry platform with study-specific user access; data can also be entered offline using a formatted spreadsheet template that can be uploaded to the database in bulk or stored locally. This ensures that data can be entered in all settings through a PC or mobile device.

The COLLECT database was described in a previous Editorial[12] aimed at scientists working in the field of pregnancy research; here we reach out to a more clinical, international readership such that we can highlight the applicability of the database to clinical research including Clinical Trial of an Investigational Medicinal Product (CTIMP) studies. The database has been designed to retain the functionality of the original SCOPE database with central data-monitoring facilities. These enable a study coordinator to monitor data entry and raise data queries with a comprehensive audit trail of data quality, which is essential in order to adhere to regulatory requirements. We have recently tested the suitability of COLLECT within these frameworks and demonstrated that the database conforms and adheres to regulations, as set out by the UK Medicines and Healthcare products Regulatory Agency (MHRA), for data collection within a CTIMP study (approved for use within ISRCTN 12695929) and in line with European Commission Regulations.

The variables within COLLECT have been based on the minimal and optimal data sets necessary to conduct and report clinical and translational research relevant to pre-eclampsia and other placental conditions.[6] The variable list has been generated to support collaborative research that will advance our understanding of the complex syndromes such as pre-eclampsia, where large and compatible data sets are necessary to unpick the complex maternal and placental phenotypes. To date, the heterogeneity of conditions such as pre-eclampsia, and the rarity of the most severe phenotypes, have limited the conclusions of recent studies.[11,13] Despite access to large (> 5000), international prospective cohorts these studies remain underpowered to investigate potential subtypes of pre-eclampsia, exemplifying the need for much larger compatible data sets.

Although COLLECT variables have been rigorously selected, they are still extensive and their collection may be beyond the scope of some researchers in low-resource settings. So CoLab is now developing a stripped-down version of the database that will function as a registry for one-time entry containing 10–15 variables per pregnancy. Importantly, this basic version will retain the modular structure of COLLECT so that it can form the backbone of more detailed data collection for specific projects. In addition, the core system should, with specific added modules, be able to replace paper-based delivery registers to record institution-wide maternity activity, without added cost.

Inevitably, however, cost is a crucial issue. For investigators in high-income countries, use of COLLECT currently

stands at $100 per month and is significantly lower than the costs associated with the design and maintenance of a study-specific database. In order to facilitate data collection for researchers in low- and middle-income countries, COLLECT has been made available free of charge in resource-poor settings.

It is very important to note that while facilitating shared use of data across research studies, the concept of a generic database platform does not in any way compromise the ownership of data within a specific study. The data collected remains the sole property of the investigators and can only be accessed by the study team, with levels of access determined by the local researchers. Following the completion of a specific study, investigators have the option to download their complete data set and remove it from the online server and/or to leave it on the central server with access still limited to the data owner until the owner requests sharing with others. As is currently the case, appropriate ethical and legal approvals need to be in place before data can be shared between collaborators.

COLLECT is available now via The Data Manager of the Global Pregnancy Collaboration, email dickensonke2@mwri.magee.edu or Tel +1 412 641 1427.

Disclosure of interests
The COLLECT database was designed by the authors. The authors do not hold any financial interests (institutional or personal) related to the use of the COLLECT database. Completed disclosure of interest forms are available to view online as supporting information.

Contribution to authorship
The manuscript was drafted by JM, with editorial input from CR, JR and LM.

Details of ethical approval
Not applicable

Funding
The Global Pregnancy Collaboration is funded by the Bill & Melinda Gates Foundation.

Acknowledgements
Not applicable. ■

References

1 Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
2 Ioannidis JP. Why most clinical research is not useful. *PLoS Med* 2016;13:e1002049.
3 Roberts JM, Mascalzoni D, Ness RB, Poston L. Collaboration to understand complex diseases: preeclampsia and adverse pregnancy outcomes. *Hypertension* 2016;67:681–7.
4 Burke O, Benton S, Szafranski P, von Dadelszen P, Buhimschi SC, Cetin I, et al. Extending the scope of pooled analyses of individual patient biomarker data from heterogeneous laboratory platforms and cohorts using merging algorithms. *Pregnancy Hypertens* 2016;6:53–9.
5 Allotey J, Snell K, Hooper R, Dodds J, Khan K, Poston L, et al. External validation, update and development of prediction models for preeclampsia using an Individual Participant Data (IPD) meta-analysis: the International Prediction of Pregnancy Complication Network (IPPIC pre-eclampsia) protocol. *Diag Prog Res* 2017;1:1–13.
6 Myatt L, Redman CW, Staff AC, Hansson S, Wilson ML, Laivuori H, et al. Strategy for standardization of preeclampsia research study design. *Hypertension* 2014;63:1293–301.
7 Burton GJ, Sebire NJ, Myatt L, Tannetta D, Wang YL, Sadovsky Y, et al. Optimising sample collection for placental research. *Placenta* 2014;35:9–22.
8 Khan K. The CROWN Initiative: journal editors invite researchers to develop core outcomes in women's health. *BJOG* 2016;123(Suppl 3):103–4.
9 Duffy JM, van 'HJ, Gale C, Brown M, Grobman W, Fitzpatrick R, et al. A protocol for developing, disseminating, and implementing a core outcome set for pre-eclampsia. *Pregnancy Hypertens* 2016;6:274–8.
10 Chappell LC, Seed PT, Myers J, Taylor RS, Kenny LC, Dekker GA, et al. Exploration and confirmation of factors associated with uncomplicated pregnancy in nulliparous women: prospective cohort study. *BMJ* 2013;347:f6398.
11 Kenny LC, Black MA, Poston L, Taylor R, Myers JE, Baker PN, et al. Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. *Hypertension* 2014;64:644–52.
12 Myatt L, Roberts JM, Redman CWG. Availability of COLLECT, a database for pregnancy and placental research studies worldwide. *Placenta* 2017;57:223–4.
13 Myatt L, Clifton RG, Roberts JM, Spong CY, Hauth JC, Varner MW, et al. First-trimester prediction of preeclampsia in nulliparous women at low risk. *Obstet Gynecol* 2012;119:1234–42.

