PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Timely access to Trial Data in the Context of a Pandemic – The Time is Now
AUTHORS	Li, Rebecca; Wood, Julie; Baskaran, Amrutha; Neumann, Stanley; Graham, Elizabeth; Levenstein, Marcia; Sim, Ida

VERSION 1 – REVIEW

REVIEWER	Georgina Humphreys Wellcome Trust, UK I am the Secretariat for the Wellcome Independent Review Panel used by many of Vivli's members.
REVIEW RETURNED	08-May-2020

GENERAL COMMENTS	Overall this is a clear and well written piece which summarises the experience of the Vivli data sharing platform to date. I have added some minor edits and comments in the attached file which I hope are useful.

REVIEWER	James McCaw University of Melbourne, Australia
REVIEW RETURNED	09-Jun-2020

GENERAL COMMENTS	The article reports on Vivli, a data sharing platform for clinical trial data, reporting on its performance characteristics in terms of data entry and data access for those seeking to use the platform. The content is clear enough. My only comment is to encourage the authors, and Vivli itself, to consider how its model fits into the growing push - with good reason in my view - towards open data, not just data sharing. There are huge opportunities with open data, but of course also challenges around (de- and re-) identification, sensitive data, privacy, compliance with regulations and research ethics etcetera. How Vivli's model for data sharing, but not open data per se fits into these emerging discussions, particularly in the light of recent high-profile issues around lack of access to source data, are important considerations. The manuscript would benefit from a rigorous engagement with those emerging discussions.

REVIEWER	Adam Dunn The University of Sydney, Australia
REVIEW RETURNED	06-Aug-2020

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript, which is a "communication" manuscript describing progress with Vivli. Noting that there does not appear to be a conflict of interest statement in the manuscript, I am reviewing this under the assumption that it would be disclosed. It describes Vivli and its progress. The following comments are mostly requests for clarity and context that might help readers who are likely to be unfamiliar with the ecosystem of trial data sharing. Introduction: It might help to define what "trial data" mean here as early as possible to help readers visualise. Readers won't understand that this is about IPD in normalised and aggregatable formats rather than structured summary results, or just publishing the summary results as tables and figures in articles. Introduction: It would also be useful to explain how rare IPD sharing is. Otherwise readers might think "5000 trials?" that's not much there are tens of thousands with summary results in CT.gov, and hundreds of thousands of articles with supplementary data published and available via PubMed and PMC. Introduction: Following from above, it would be nice to place this in better context relative to YODA and other initiatives aimed at sharing clinical trial data at IPD or summary levels, including when they started, their size, and what they do or don't provide. It's really hard to see the value in the current manuscript without "triangulating" the importance of Vivli. Introduction: One of the most interesting aspects of Vivli and what sets it apart from other initiatives (though there are obvious similarities with YODA and others) is the governance model, and the process for providing access via requests. Perhaps some more detail on that process would be useful - as it stands the information is a bit opaque and it is hard to see how it works. Introduction: How those data are connected and then used is not clear. It sounds good that there are ways to combine data from multiple sources but how does that actually work in practice? Why does it matter? How many of the requests have taken advantage of data from other data sources? Introduction: The COVID-19 introduction paragraph and the brief description of the COVID-19 portal aren't fully described. It's not clear what benefit readers will get from knowing that it exists. Data access and metrics: It would be nice to have a more detailed graphic representation of the number of requests over time; perhaps a stacked bar chart with the number that were requested per month (?), coloured by the number that were rejected, fulfilled, in progress, etc. Figure 1 absolutely should not be a bar chart with the cumulative requests - an inappropriate representation for cumulative data. Data access and metrics: It would also be nice to see how many of the 5000 trials have been requested (among the 125 requests have some been requested many times, and many none?) as well as how often access to certain trials has been granted (any that have been requested many times and always rejected?). Perhaps a graphic?

	Key insights: IPD is introduced for the first time here as an acronym. As above, it should be introduced much earlier. Key insights: This is perhaps the most interesting part of the manuscript, and this section could be expanded to explain how these insights were determined, and to provide an example or evidence to back them up.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name
Georgina Humphreys
Institution and Country
Wellcome Trust, UK

We have considered the useful feedback provided by this reviewer and updated the paper and made the additions and revisions that were feasible.

Reviewer: 2
Reviewer Name
James McCaw

Institution and Country
University of Melbourne, Australia

Please state any competing interests or state 'None declared':
None declared

Please leave your comments for the authors below
The article reports on Vivli, a data sharing platform for clinical trial data, reporting on its performance characteristics in terms of data entry and data access for those seeking to use the platform.

The content is clear enough. My only comment is to encourage the authors, and Vivli itself, to consider how its model fitsnto the growing push - with good reason in my view - towards open data, not just data sharing.

There are huge opportunities with open data, but of course also challenges around (de- and re-) identification, sensitive data, privacy, compliance with regulations and research ethics etcetera. How Vivli's model for data sharing, but not open data per se fits into these emerging discussions, particularly in the light of recent high-profile issues around lack of access to source data, are important considerations.
Thank you for this helpful and timely comment. We agree this is an important aspect and there is a spectrum of data sharing models ranging from open to more restricted data sharing and certainly a more recent push towards more open data sharing. Vivli has had a philosophy to meet data contributors "where they are" to encourage the broadest participation among a diverse array of stakeholders which has meant a "managed access" approach. We have discussed this approach under the "Objectives" section.
Reviewer: 3
Reviewer Name

Adam Dunn

Institution and Country
The University of Sydney, Australia

Please state any competing interests or state ‘None declared’:
None declared

Please leave your comments for the authors below
Thank you for the opportunity to review this manuscript, which is a "communication" manuscript describing progress with Vivli. Noting that there does not appear to be a conflict of interest statement in the manuscript, I am reviewing this under the assumption that it would be disclosed. It describes Vivli and its progress. The following comments are mostly requests for clarity and context that might help readers who are likely to be unfamiliar with the ecosystem of trial data sharing.

Our competing interests are added in the BMJ system and will appear as below.

Vivli receives funding support from the Arnold Ventures, Doris Duke Charitable Trust, the Helmsley Charitable Trust, ph

Introduction: It might help to define what "trial data" mean here as early as possible to help readers visualise. Readers won't understand that this is about IPD
in normalised and aggregatable formats rather than structured summary results, or just publishing the summary results as tables and figures in articles.

Thank you for this excellent point. We have added several sentences to the beginning of the abstract to inform the reader understand the uniqueness of IPD data and how IPD or the “raw data” from each participant is used and that historically the aggregate or summary-level data is shared via publications.

Introduction: It would also be useful to explain how rare IPD sharing is. Otherwise readers might think "5000 trials?" that's not much there are tens of thousands with summary results in CT.gov, and hundreds of thousands of articles with supplementary data published and available via PubMed and PMC.
We have expanded this point in the Objective section. This is now the first sentence of the abstract as well.

Introduction: Following from above, it would be nice to place this in better context relative to YODA and other initiatives aimed at sharing clinical trial data at IPD or summary levels, including when they started, their size, and what they do or don't provide. It's really hard to see the value in the current manuscript without "triangulating" the importance of Vivli.
Excellent point, We inserted a sentence in the second paragraph under the objective section explaining that Vivli was designed to bridge the fragmentation in the current data sharing ecosystem and as such is bridging platforms such as YODA, SOAR, government platforms and others.

Introduction: One of the most interesting aspects of Vivli and what sets it apart from other initiatives (though there are obvious similarities with YODA and others) is the governance model, and the process for providing access via requests. Perhaps some more detail on that process would be useful - as it stands the information is a bit opaque and it is hard to see how it works.
We have split the Data Access process section from the Data access metrics section and inserted a new “figure 1” that illustrates the detail on the process for providing access to data.

Introduction: How those data are connected and then used is not clear. It sounds good that there are

ways to combine data from multiple sources but how does that actually work in practice? Why does it matter? How many of the requests have taken advantage of data from other data sources?

We have re-organized this section to emphasize that 25% of the requests involve combining data from multiple sources and discuss how this occurs in practice as well as "why this matters" scientifically. If Vivli did not exist, the ability to integrate these data from multiple sources would not be possible in most cases. For example, we have enabled datasets to be combined and accessed from multiple sources in ways not previously possible, by connecting with existing platforms (such as the YODA Project, ImmPort, and NHLBI's BioLINCC and enabling the combination from diverse data sources (academic institutions, large pharma, NIH and small biotech).

Introduction: The COVID-19 introduction paragraph and the brief description of the COVID-19 portal aren't fully described. It's not clear what benefit readers will get from knowing that it exists.

We have elaborated the need for the COVID portal from a researcher's perspective and how the portal facilitates sharing and access. Particularly in the area of anonymization where this is a primary barrier to sharing.

Data access and metrics: It would be nice to have a more detailed graphic representation of the number of requests over time; perhaps a stacked bar chart with the number that were requested per month (?), coloured by the number that were rejected, fulfilled, in progress, etc. Figure 1 absolutely should not be a bar chart with the cumulative requests - an inappropriate representation for cumulative data.

We have reconfigured this figure in a line graph representative of change over time and agree the bar graph was not appropriate representing the cumulative data requests. We have indicated the overall number of rejected, withdrawn and fulfilled.

Data access and metrics: It would also be nice to see how many of the 5000 trials have been requested (among the 125 requests have some been requested many times, and many none?) as well as how often access to certain trials has been granted (any that have been requested many times and always rejected?). Perhaps a graphic?

Thank you for this helpful suggestion. This information regarding how many of the 5,446 studies have been requested has been added under data access metrics (20.1%) or 1136 studies. We have also added information on frequently requested studies.

Key insights: IPD is introduced for the first time here as an acronym. As above, it should be introduced much earlier.
We have added an explanation of IPD data sharing to the introduction section of the abstract.
Key insights: This is perhaps the most interesting part of the manuscript, and this section could be expanded to explain how these insights were determined, and to provide an example or evidence to back them up.

This section has been expanded and we have provided several details to back up this section.

VERSION 2 – REVIEW

REVIEWER	Georgina Humphreys

	Wellcome Trust, UK I act as secretariat for the Wellcome IRP used by some of Vivli's members to assess data access requests.
REVIEW RETURNED	07-Oct-2020

GENERAL COMMENTS	I think the manuscript has been improved following revisions and is now acceptable for publication.