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)	Navigating data governance associated with real-world data for public benefit: an overview in the UK and future considerations
AUTHORS	Jones, Monica; Stone, Tony; Mason, Suzanne; Eames, Andy; Franklin, Matthew

VERSION 1 – REVIEW

REVIEWER	Harry Hochheiser University of Pittsburgh
REVIEW RETURNED	27-Dec-2022

GENERAL COMMENTS	In this brief note, Jones et al. provide a review of recent regulations regarding the use of real-world data, along with a discussion of recommendations that should be considered by those in this field. Although this short introduction is necessarily limited inn scope and discussion of nuanced details, this presentation of relevant regulations and related efforts is still valuable and informative. I have a few suggestions for enhancements that might increase the readability and utility of the work:
	1. This paper introduces several frameworks, including the GDPR, the NHS act, the COPI notices, and the Health and Care Bill, and the five safes framework. However, it's hard to see how these different regulations relate - are there common areas? Important differences? Some form of summary framework would be very helpful. Ideally, this summary might compare these regulations along some common, salient dimensions - perhaps the five safes - that might indicate which questions are included or omitted from each framework. Barring this, a table summarizing each framework would be helpful.
	2. Building on point 1, this paper is lacking in discussion of the perspectives of the interested parties: researchers, administrators, clinicians, and members of the public will each interact with these regulations in different ways. Discussion of some of these differences - or at least acknowledgment of the need to consider these perspectives might be possible. There is some acknowledgment in the penultimate paragraph of Section 7 - more might be helpful. For example, how do the responsibilities of researchers differ from those of administrators?

	3. Regarding recommendations, there are discussion of trusted research environments, but this is only one possible approach. I am admittedly much more familiar with how research is conducted in the US than in the UK, but I wonder if some discussion of other arrangements, such as pairwise collaboration between health organizations - might also be an important topic of discussion. In the US context, this is a regular occurrence, arguably more prevalent than trusted research environments. As in point #2, it might help to be more explicit about the various parties involved and which recommendations are relevant to which classes of participants.
	4. As this paper will likely be read by researchers outside of the UK, some acknowledgment of differences between jurisdictions might be helpful. HIPAA in the US context would be an obvious example.
	Two small points:
	1. The five safes concept is introduced in Section 6 without a reference. If this is not a unique contribution of the current work, the original source should be cited.
	2. There is an incomplete reference in the second paragraph of Section 7.

REVIEWER	Jiyeon Kang London School of Hygiene and Tropical Medicine Faculty of Public Health and Policy, Health Services Research and Policy
REVIEW RETURNED	22-Feb-2023

GENERAL COMMENTS	Thank you for sharing this interesting paper. This paper highlights the practical aspects of data governance. Given the current interest in real-world data (RWD), I believe this paper will provide helpful information to readers who want to understand the basic law/regulation when they access public data. Although I found this paper interesting, more elaboration will help this paper be better shaped. I included some comments on each section, which might be helpful to articulate this paper.
	1. Introduction The background should more clearly include current practical and legal issues. The introduction covers general information regarding RWD and what RWD can do. This part seems a bit lengthy. In the first part explaining terminology about the data, quite a few sentences are repetitive. For example, "Often referred to as 'administrative', …." "This includes administrative data, registry data … " These sentences can be tidied up.
	I understand this paper tries to provide an overview with respect to data governance. It is interesting to read what sort of rules need to be considered. However, it would be essential to address the practical issues with respect to data governance. For example, the practical challenges when dealing with the data due to the GDPR would be interesting and valuable for the readers.
	The last paragraph in the introduction needs to be more elaborated.

For example, "Potential public benefit is forfeited by the under-utilisation of this data for secondary uses in part due to risk aversion when faced with the prospect of navigating necessary and important data governance processes." This is simply not true from previous literature. Even in HTA decision-making, especially NICE decision-making, RWD have been widely used to help to reduce uncertainty in decision-making. Also, medical devices are more likely to open to use such data.

Moreover, while it is true that data governance is a barrier to accessing such data, I don't think that data governance is not the main reason for less using such data. I agree that data governance can be a complicating issue, which potentially affects to decrease the demand. However, the claim seems too strong. It would be better to paraphrase to avoid misinterpretation.

P. 5 Line 34
What is IG?

4. A moment of realisation
Although the notice allows more active access, is this notice universal for all cases or still valid? It might be helpful to address which cases are applied or not more clearly.

5. Instigating further change
What is the main challenge or daunting part that occurred by the Health and Care Bill?
Section 5 summarises the Bill, but the more important and interesting part to the audience might be what kind of potential challenges this Bill brings.

7. Suggestion
It would be great if the examples (HDR UK, YHCR) were provided with more information. Although this paper is not the paper about this dataset, it would be interesting to read if there are any different approaches to overcome the data governance issue.

"Enabling access to data for secondary uses through TREs complying with the Five Safes Framework reassures data suppliers and the public compared to provisioning data extracts for individual projects[]."
The reference is missing.

I agree that this kind of initiative or strategy is required to use collected data more efficiently! It might be worth addressing NHS's attempt to connect the datasets across the NHS trust as a part of an investment.

While it is a minor point, using data in the plural would be better.

VERSION 1 – AUTHOR RESPONSE

No.	Reviewers' comments: reviewer #2	Authors' responses	Section
0	Thank you for sharing this interesting paper. This paper highlights the practical aspects of data governance. Given the current interest in real-world data (RWD), I believe this paper will provide	We thank the reviewer for their comments	N/A

	helpful information to readers who want to understand the basic law/regulation when they access public data. Although I found this paper interesting, more elaboration will help this paper be better shaped.		
1	Introduction: The background should more clearly include current practical and legal issues. The introduction covers general information regarding RWD and what RWD can do. This part seems a bit lengthy. In the first part explaining terminology about the data, quite a few sentences are repetitive. For example, "Often referred to as 'administrative', …." "This includes administrative data, registry data … " These sentences can be tidied up.	We have shortened the introduction in terms of general information about RWD, to make it more concise. We have kept the introduction as being light in terms of detail about current practical and legal issues, given this will be the focus of the other sections of the article and so described throughout.	Section 1
2	Introduction: I understand this paper tries to provide an overview with respect to data governance. It is interesting to read what sort of rules need to be considered. However, it would be essential to address the practical issues with respect to data governance. For example, the practical challenges when dealing with the data due to the GDPR would be interesting and valuable for the readers.	Practical issues are in-part described in Section 7 "What else is needed: a change in the data landscape" and Section 9 "Practical example of approach to data governance". As the article focuses on information governance and legal aspects for consideration, practical steps (e.g. transfer of data, anonymisation) are aspects for consideration; however, we have chosen not to go into too much detail in this regard so we can focus more on the information governance and legal aspects (i.e. as the pertinent purview of the paper) and to avoid the article becoming too long due to trying to cover other related, but also complicated aspects for consideration such as the range of practical challenges.	Section 7 Section 9
3	Introduction: The last paragraph in the introduction needs to be more elaborated. For example, "Potential public benefit is forfeited by the under-utilisation of this data for secondary uses in part due to risk aversion when faced with the prospect of navigating necessary and important data governance processes." This is simply not true from previous literature. Even in HTA decision-making, especially NICE decision-making, RWD have been widely used to help to reduce uncertainty	We don't wholly agree with the reviewer's comment, while also noting that there is publication bias in terms of reporting under-utilisation of RWD for public benefit. For example, researchers don't tend to publish when they can't get data to do research; they typically more often publish when they have managed to get the data for research. Thus it isn't wholly known how often research leading to public benefit hasn't happened due to risk aversion to dealing with information governance, but we can use qualitative methods to better understand such barriers; e.g. see our qualitative study: http://dx.doi.org/10.1186/s12889-023-15035-w	Section 1

	in decision-making. Also, medical devices are more likely to open to use such data.

Moreover, while it is true that data governance is a barrier to accessing such data, I don't think that data governance is not the main reason for less using such data. I agree that data governance can be a complicating issue, which potentially affects to decrease the demand. However, the claim seems too strong. It would be better to paraphrase to avoid misinterpretation. | We agree that data governance is perhaps not the main and only reason for RWD under-utilisation, but it is suggested as a reason. We conducted a qualitative study which included stakeholders perspectives (i.e. local government agents, clinicians, data analysts, researchers, and the public), within which suggested barriers to using and sharing RWD included governmental and legal, organisational features, and process factors which adversely affect the sharing of real world data (http://dx.doi.org/10.1186/s12889-023-15035-w).

We have changed the sentence to now read as:

"Potential public benefit is forfeited by the under-utilisation of this data for secondary uses, in part due to risk aversion when faced with the prospect of navigating necessary and important data governance processes, although there are other barriers to sharing and using RWD which can include organisational features and process factors (REF)" *The REF is this paper: http://dx.doi.org/10.1186/s12889-023-15035-w | |
| 4 | P. 5 Line 34. What is IG? | IG refers to Information Governance. We have added a new Section, Section 2, which clarifies the acronym and briefly answers the question "What is information governance (IG)?" | Section 2 |
| 5 | A moment of realisation: Although the notice allows more active access, is this notice universal for all cases or still valid? It might be helpful to address which cases are applied or not more clearly. | We have now further clarified that the COPI Notices to which we refer were specific to managing the COVID-19 pandemic (the specifics around the remit of the notices are in the references provided), as well as that most notices expired as of 30th June 2022. | Section 5 |
| 6 | Instigating further change: What is the main challenge or daunting part that occurred by the Health and Care Bill?

Section 5 summarises the Bill, but the more important and interesting part to the audience might be what kind of potential challenges this Bill | We have now changed this section to refer to the Health and Care Act 2022, rather than the Bill which preceded it.

We describe the main challenge as being is to implement significant changes to the organization and operation of the National Health Service (NHS), including related to accountability and governance. | Section 6 |

	brings.		
7	It would be great if the examples (HDR UK, YHCR) were provided with more information. Although this paper is not the paper about this dataset, it would be interesting to read if there are any different approaches to overcome the data governance issue.	We now describe HDR UK and YHCR in more detail, in terms of practical examples of approaches to data governance as a pertinent aspect for consideration within our article – this is included in our new Section 9 "Practical examples of approaches to data governance"	Section 9
8	"Enabling access to data for secondary uses through TREs complying with the Five Safes Framework reassures data suppliers and the public compared to provisioning data extracts for individual projects[]." The reference is missing.	A Five Safes reference has now been added	References Section 10
	I agree that this kind of initiative or strategy is required to use collected data more efficiently! It might be worth addressing NHS's attempt to connect the datasets across the NHS trust as a part of an investment.	We now have a new Section 10 "How the NHS connects datasets as an investment", which refers to the initiative (and associated investment) of integrating various datasets within the NHS to create a comprehensive and interoperable health information system to improve patient care, enhance research capabilities, and drive efficiencies within the healthcare system.	
9	While it is a minor point, using data in the plural would be better.	We have screened the article to check and change as appropriate, the use of data in the plural	Throughout

VERSION 2 – REVIEW

REVIEWER	Harry Hochheiser University of Pittsburgh
REVIEW RETURNED	10-Jun-2023

GENERAL COMMENTS	Thanks to the reviewers for their thoughtful revisions to the paper. The additions - particularly the comparison to the US context - are particularly helpful, and the paper is much improved. The discussion of the five Safes is particularly helpful.
	That said, there are a few lingering issues that should be addressed.
	1. The framing of the problems is still very UK-centric. I suggest retaning the mention of the UK in the title of the paper.
	2. The discussion in section has been enhanced, but it ss still confusing. Both pairwise collaboration and TREs are not particularly well-defined. Who runs TREs, how do they work, and what does "pairwises collaboration in the context of a TRE" mean"? This is unclear. This problem is particularly acute with respect to Table 4,

	as it's not clear how these advantages accure to pairwise collaboration in the context of TREs vs. any other use of TREs.

REVIEWER	Jiyeon Kang London School of Hygiene and Tropical Medicine Faculty of Public Health and Policy, Health Services Research and Policy
REVIEW RETURNED	01-Aug-2023

GENERAL COMMENTS	Thank you for sharing your revised manuscript. Overall, this version is more precise than the previous one. There are a couple of comments, which are pretty minor this time. In "2. What is information governance (IG)?", it would be better to include the reference to avoid plagiarism and increase credibility. Also, if the abbreviation, IG, is decided to be used, then using it in the whole context would be easier for an audience to understand. Currently, it is mixed: the abbreviation is used in some parts, whereas in others, it is not. Another minor point is about Clinical Practice Research Datalink (CPRD). Although this paper discussed linking various databases not only collected in the clinical practice, CPRD is one of the most extensive linked databases with great potential. GP data, HES data, ONS data and SACT data are linked in this database, reducing hassles and giving more research opportunities. It might be worth mentioning that there has been an effort to combine this data at the national level, but still, there are legal challenges, as the paper described.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1: Dr. Harry Hochheiser, University of Pittsburgh

1. We will provide a summary framework as suggested comparing against common dimensions i.e. the Five safes

2. Although originally this paper was intended as summary of the information governance regulations to aid researchers to navigate this field we will aim to provide a discussion of the differences between them

3. The direction of travel in the UK is strongly towards TRE/SDEs and is government led although there is still some merit in pairwise collaboration whilst this is put in place

4. We will aim to acknowledge the differences between jurisdictions i.e. HIPAA

NB. updates to references will be made as suggested.

Reviewer 2: Ms.Jiyeon Kang, London School of Hygiene and Tropical Medicine Faculty of Public Health and Policy

1. We will update the background to cover current practice and legal issues as suggested and include practical examples of challenges 2. Paraphrase claims to avoid misinterpretation

3. Expand Information Governance (IG)

4. We will aim to address the cases that this applies to

5. State the main challenge of the Health and Care Bill

6. No text

7. Add a little more detail about HDRUK and YHCR ref practical examples of approach to data governance, add references and address the NHS attempt to connect datasets as an investment. Consider plural.

VERSION 3 – REVIEW

REVIEWER	Harry Hochheiser University of Pittsburgh
REVIEW RETURNED	04-Sep-2023

GENERAL COMMENTS	Thanks for your responses to previous comments.