PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The comprehensive evaluation of a prospective Australian patient cohort with suspected genetic kidney disease undergoing clinical genomic testing- a study protocol
AUTHORS	Jayasinghe, Kushani; Stark, Zornitza; Patel, Chirag; Mallawaarachchi, Amali; McCarthy, Hugh; Faull, Randall; Chakera, Aron; Sundaram, Madhivanan; Jose, Matthew; Kerr, Peter; Wu, You; Wardrop, Louise; Goranitis, I; Best, Stephanie; Martyn, Melissa; Quinlan, Catherine; Mallett, Andrew

VERSION 1 – REVIEW

REVIEWER	Francois Rousseau MD MSc FRCPC FCAHS Université Laval, Canada
REVIEW RETURNED	26-Feb-2019

GENERAL COMMENTS	The major bias of this study is that patient's diagnostic outcome after NGS testing will be compared with the patient'S outcome before this genomic investigation. Thus it seems obvious that the comparator will have an inferior performance than with this extra test. Usually, in the context of absence of strong evidence of clinical utility (i.e. improved patient health outcomes), it is possible to design a RCT to minimize biases. The argument proposed here for not doing so does not appear so compelling if there is no study showing improved patient outcomes with NGS (which is in the rationale for doing this study). The tests NGS that will be used are very different in their potential yield and it would have been preferable to study one of them specifically, otherwise the power of the study will be weakened by the analysis of subgroups. The protocol mentions that only the genes of interest will be analyzed and reported. Given the number of centers it would have been preferable to have a predefined set of genes agreed upon between centers for each phenotype presentation. A one year time horizon for the economic model appears somewhat insufficient especially for the study of QALYs. Most treatments for genetic conditions to be identified will likely take more than 12 months to provide results and long term benefits may be more important than short term ones. It would have been good to specify that these cost effectiveness and cost utility analyses will follow some national or international guidelines for economic studies. There are also many missing elements in the description of the Health economic evaluation. How will the unit costs be estimated ? How will the modeled decision trees be validated ? Will there be sensitivity analyses ? Etc. The 14-lines

	description provided is insufficient to be considered a complete protocol for a health economics evaluation. Clinical utility is usually defined as improved patient outcomes (NASEM - An evidence framework for genetic testing, 2017). Clinical utility is usually not considered equivalent to « clinical usefulness ». It appears that the authors consider clinical utility as a change in clinical investigations and management of patients as opposed to improved patient health outcomes. This is not the usual meaning of clinical utility and it would appear preferable to use the term « clinical management ». I have not found in the MS justification for the sample size, nor power calculations based on estimated differences and standard deviations for the primary outcome. Given that the authors have a first sample, such calculations should have been possible.

REVIEWER	Kurt Christensen Brigham and Women's Hospital and Harvard Medical School, USA
REVIEW RETURNED	25-Apr-2019

GENERAL COMMENTS	Comments on bmjopen-2019-029541 “A study protocol for comprehensive evaluation of clinical genomic testing in patients with suspected genetic kidney disease” Manuscript bmjopen-2019-029541 summarizes a protocol to evaluate a multi-center genomic testing program for patients with suspected genetic kidney disease. The manuscript is written well, and the program and evaluation are appropriate. My most important concern is that the authors have not explained the generalizable knowledge for this manuscript. What lessons can other studies learn from this group’s process for developing a protocol (said another way, why not simply register the study protocol, where readers can see modifications that were incorporated over time). I am confident that the authors can address this concern and my others fairly easily. Other major concerns In the Data Analysis section, the authors write, “Results on diagnostic and clinical utility will be expressed as frequencies and percentages for categorical variables and be compared using the chi-square…” The authors present the aim as the proportion of patients with suspected GKD who obtain a positive diagnosis following genomic testing compared with their clinical diagnosis prior to testing. Given the inclusion/exclusion criteria, won’t zero patients have a clinical diagnosis have a clinical diagnosis prior to testing? I was struggling to understand how the replication cohort would be analyzed. Presumably, since many of them do have a molecular diagnosis (and maybe were in the baseline cohort?), they’ll just be evaluated for the outcomes listed on the evaluation surveys? The rationale and analytic approach for the replication cohort should be clarified. It would be helpful to understand why patients who undergo only Sanger sequencing are excluded from the study. I think the economic analyses would benefit from a little elaboration. How are findings from the observational research – which includes instruments that assess health-related quality of life and methods to estimate health care utilization and health care costs – being integrated into the decision model?

	Minor Issues In the Strengths and Limitations section, the statement, “Patients will be recruited for clinically indicated genomic testing for suspected genetic kidney disorders,” seems neither a strength nor a limitation. The Aim, “To determine the clinical impact of genomic testing at three months follow-up in a cohort” seems like it has a typo. Should it read, “to determine the clinical impact of genomic testing after three months in a cohort…” Page 6, line 7: define “CKD” Page 8, line 13: “Patients with an existing molecularly confirmed genetic diagnosis” Page 9, line 24: “Participants are asked to complete the first survey at 2-3 weeks after attending the RGC.” What happens if they have multiple visits to the RGC? For the measurement of decision regret, what decision is being analyzed? The decision to visit an RGC? The authors write, “Options to consent to accessing the Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits scheme (PBS), hospital and emergency data sets and data sharing in addition to study evaluation data collection.” It would be helpful to very briefly explain these resources for an international audience that is unfamiliar with the Australian health system. The authors use the term, “relative cost-effectiveness.” I suggest changing “relative” to “incremental” The abstract says that the expected cohort is 360 patients, but later, the authors write that the target sample size is 500 patients across both the baseline and replication cohorts. In the Strengths and Limitations, the authors wrote, “Due to the heterogeneity of genetic testing availability in Australia, it will not be equitable or practical to have a concurrent control arm for comparison.” Later, they wrote, “Due to the rare nature of GKD, and because genetic testing is performed as part of routine clinical care in most states/centres, it would not be equitable to have a control arm for this study.” I was unconvinced by either of these rationales and didn’t follow their logic. That said, the authors provide a convincing and coherent rationale in the beginning of the Discussion section (i.e., because genetic testing is clinically indicated and is standard-of-care, it would be unethical to deny genetic testing). I found the Participant Identifiers section to be a little hard to follow. Aren’t most of the investigators involved with the clinical care of the patients?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Francois Rousseau MD MSc FRCPC FCAHS
Institution and Country: Université Laval, Canada
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

The major bias of this study is that patient's diagnostic outcome after NGS testing will be compared with the patient'S outcome before this genomic investigation. Thus it seems obvious that the

comparator will have an inferior performance than with this extra test. Usually, in the context of absence of strong evidence of clinical utility (i.e. improved patient health outcomes), it is possible to design a RCT to minimize biases. The argument proposed here for not doing so does not appear so compelling if there is no study showing improved patient outcomes with NGS (which is in the rationale for doing this study).
Thank you for your comment, we understand that the argument proposed for not pursuing an RCT should be explained better and have amended the points made in the 'strengths and limitations', 'methods' and 'discussion' section. The main reason we have not pursued an RCT is because genomic testing is used in many instances and is offered already in several states. Therefore, it would not be ethical to deny some patients of testing for the purpose of this study. During the development of the initial protocol, we aimed to also investigate those who did not undergo genomic testing. However, the initial formal and informal community and consumer review feedback was that it was unacceptable to not offer testing to some individuals, and patients were not willing to take part in the study unless they were offered genomic testing. Furthermore, the genetic counsellors also felt it was unethical for patients to take their time attending clinics and competing surveys, especially if it represented structure denial of a test they would otherwise undergo in our current clinical context.

The tests NGS that will be used are very different in their potential yield and it would have been preferable to study one of them specifically, otherwise the power of the study will be weakened by the analysis of subgroups.
Thankyou for pointing out this important difference. It is indeed true that various NGS tests are present. The reason we wish to study genomic testing (as opposed to panel exome-based testing, whole exome or whole genome sequencing only) is that this is the existing state of testing available in our resource context. It is important to note that the baseline requirements of inclusion of tests are they all need to be multigene panels or genomic tests, and hence single gene testing is not included. Therefore, only patients undergoing genomic tests in clinically accredited laboratories will be included. We agree that there if potential heterogeneity between all the different approaches but given that this is a pragmatic research study, we feel it is important to determine the implications in this context. Please also note that our main analysis does not focus on analysing subgroups by the type of NGS, but rather by other clinical factors.

The protocol mentions that only the genes of interest will be analyzed and reported. Given the number of centers it would have been preferable to have a predefined set of genes agreed upon between centers for each phenotype presentation.

Thank you for pointing this out. We agree entirely that it is preferable to have the set of genes that will be investigated with relation to kidney disease predefined and had established a clinical system of kidney disease genes for this study. This list has since been refined and published, and we have referenced this in the paper accordingly(1).
A one year time horizon for the economic model appears somewhat insufficient especially for the study of QALYs. Most treatments for genetic conditions to be identified will likely take more than 12 months to provide results and long term benefits may be more important than short term ones. It would have been good to specify that these cost effectiveness and cost utility analyses will follow some national or international guidelines for economic studies. There are also many missing elements in the description of the Health economic evaluation. How will the unit costs be estimated ? How will the modeled decision trees be validated ? Will there be sensitivity analyses ? Etc. The 14-lines description provided is insufficient to be considered a complete protocol for a health economics evaluation.

Thank you for the opportunity to revise our economic evaluation plans in light of these comments. The section has now been updated to include more specific information around the cost-effectiveness analysis (page 10).
Clinical utility is usually defined as improved patient outcomes (NASEM - An evidence framework for genetic testing, 2017). Clinical utility is usually not considered equivalent to « clinical usefulness ». It appears that the authors consider clinical utility as a change in clinical investigations and management of patients as opposed to improved patient health outcomes. This is not the usual meaning of clinical utility and it would appear preferable to use the term « clinical management ».

Thankyou for clarifying this point of difference. The current literature defines 'utility' as broader term in the context of genomics(2-4). We feel that the definition of 'clinical utility' as restricted to changes in health outcomes alone is insufficient for genomic medicine. Instead, we define clinical utility as improved health outcomes following a test result, reflected by changes in patient behaviour and clinical decisions. This definition is also used in many systematic reviews focussing on clinical genomics(5) However, we contend that the term clinical utility represents more than just changes in clinical management and have therefore amended the use of this term and changed this to 'diagnostic utility and clinical usefulness'.

I have not found in the MS justification for the sample size, nor power calculations based on estimated differences and standard deviations for the primary outcome. Given that the authors have a first sample, such calculations should have been possible.
Thankyou for your comment. While we have a previous first sample, we believe there is insufficient published data to calculate a sample size for this study. It would have been possible to power this study to detect a 10% change in the diagnostic rate between the clinical and genomic diagnostic methods, however the diagnostic utility is only a small part of the possible benefit of genomic testing. It is also important to also capture the clinical implications and personal utility of genomic testing, which we cannot power for. Therefore we have deliberately adopted a descriptive design in order to capture the potential benefits of genomic testing.
Comments on bmjopen-2019-029541
"A study protocol for comprehensive evaluation of clinical genomic testing in patients with suspected genetic kidney disease"
Manuscript bmjopen-2019-029541 summarizes a protocol to evaluate a multi-center genomic testing program for patients with suspected genetic kidney disease. The manuscript is written well, and the program and evaluation are appropriate. My most important concern is that the authors have not explained the generalizable knowledge for this manuscript. What lessons can other studies learn from this group's process for developing a protocol (said another way, why not simply register the study protocol, where readers can see modifications that were incorporated over time). I am confident that the authors can address this concern and my others fairly easily.

Thankyou for your comment. We believe it is important for readers to be able to view this protocol in published format for several reasons. Firstly, it will be a detailed protocol available for all nephrologists and other relevant clinicians in Australia to access, which will enable them to reliably recruit patients into this study. More importantly, we wish to highlight the this is the first study to sequence patients with kidney disease in a pragmatic, clinical setting which uses treating nephrologists as the referral base. Hence these results are more likely to be replicated in a clinical environment and be more generalisable to practising clinicians who wish to make informed management decisions. By publishing this protocol, we hope to improve access to genomic testing using a generalisable broad population throughout Australia. Finally, to our knowledge, this is the first time that a structure, objective and unbiased approach is taken to measure the outcomes of genomic testing in kidney disease, which we believe is more useful as a publication as opposed to just registration. We have highlighted this more clearly in the discussion section of the manuscript.

Other major concerns
In the Data Analysis section, the authors write, “Results on diagnostic and clinical utility will be expressed as frequencies and percentages for categorical variables and be compared using the chi-square…” The authors present the aim as the proportion of patients with suspected GKD who obtain a positive diagnosis following genomic testing compared with their clinical diagnosis prior to testing. Given the inclusion/exclusion criteria, won’t zero patients have a clinical diagnosis have a clinical diagnosis prior to testing?

Thank you for focussing our attention to this clarification in the methods. The primary outcome will be the proportion of patients with a positive molecular diagnosis following genomic testing (ie the diagnostic rate or yield). The secondary outcome is indeed the proportion of patients with a change in clinical diagnosis following genomic testing. Patients will be included if their clinicians wish to clarify or confirm an existing clinical diagnosis and/or mode of inheritance or are unsure between multiple differential diagnoses, or if they are unsure of a diagnosis altogether. Hence the pre-test differential diagnosis will be compared to the post-test genomic diagnosis. We have amended the methods section to clarify the indications for referring for testing (ie inclusion criteria).

I was struggling to understand how the replication cohort would be analyzed. Presumably, since many of them do have a molecular diagnosis (and maybe were in the baseline cohort?), they’ll just be evaluated for the outcomes listed on the evaluation surveys? The rationale and analytic approach for the replication cohort should be clarified. It would be helpful to understand why patients who undergo only Sanger sequencing are excluded from the study.

Thank you for this query. We have clarified that the replication cohort is mutually exclusive from the baseline cohort, ie, a participant can be included in one but not both of the baseline and prospective cohorts. Further, whilst the replication cohort participants will not undertake evaluation surveys, their outcomes will be assessed against all other measures in common with the baseline cohort. The rationale of this is that given the participants will often have rare or ultra-rare kidney diseases, we anticipate a wide variety of patient, disease and sub-cohort level outcomes to become apparent as the overall cohort is analysed at larger scale over a longer observation timespan. Those undergoing Sanger sequencing alone have been excluded for two main reasons. Firstly, this would not represent a truly genomic approach to diagnosis and we specifically are attempting to assess the impact and utility of such newer multi-gene clinical genomic sequencing approaches in clinical practice rather than existing single-gene Sanger sequencing approaches. Secondly, it can be difficult to ascertain whether or not sole Sanger sequencing might represent primary, segregation or cascade testing for an individual patient, and so this would add confusion and limitation to outcome assessment in terms of whether the diagnostic outcome is one which is newly identified or simply a previously known familial variant/diagnosis. We believe that these are all important points to clarify and have added them to the manuscript.

I think the economic analyses would benefit from a little elaboration. How are findings from the observational research – which includes instruments that assess health-related quality of life and methods to estimate health care utilization and health care costs – being integrated into the decision model?

Thank you for this comment. In response to this question and the comment received from Reviewer 1, the economic evaluation section has been updated (page 10).

Minor Issues
In the Strengths and Limitations section, the statement, “Patients will be recruited for clinically indicated genomic testing for suspected genetic kidney disorders,” seems neither a strength nor a limitation.

Thank you. We have elaborated on this point in the manuscript. We feel that the fact that patients are undergoing genomic testing in a clinical rather than a research context, and that this pragmatic approach therefore represents a strength as the results are likely to be more generalisable for clinicians.

The Aim, “To determine the clinical impact of genomic testing at three months follow-up in a cohort” seems like it has a typo. Should it read, “to determine the clinical impact of genomic testing after three months in a cohort…”

Thank you, have amended.

Page 6, line 7: define “CKD”
Thank you, have amended

Page 8, line 13: “Patients with an existing molecularly confirmed genetic diagnosis”
Thank you, have amended.

Page 9, line 24: “Participants are asked to complete the first survey at 2-3 weeks after attending the RGC.” What happens if they have multiple visits to the RGC?

The purpose of this is to measure their experience post the RGC prior to receiving results. They will still only complete two surveys in total regardless of the number of visits to the RGC.

For the measurement of decision regret, what decision is being analyzed? The decision to visit a RGC?

Thank you, the decision being analysed here is to undergo genomic testing. We have amended the section under ‘evaluation surveys’.

The authors write, “Options to consent to accessing the Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits scheme (PBS), hospital and emergency data sets and data sharing in addition to study evaluation data collection.” It would be helpful to very briefly explain these resources for an international audience that is unfamiliar with the Australian health system.

Thank you. We appreciate this valid point and have detailed this in the manuscript under the heading ‘sequencing’

The authors use the term, “relative cost-effectiveness.” I suggest changing “relative” to “incremental”

Thank you, have amended

The abstract says that the expected cohort is 360 patients, but later, the authors write that the target sample size is 500 patients across both the baseline and replication cohorts.

Thank you, have amended the abstract accordingly

In the Strengths and Limitations, the authors wrote, “Due to the heterogeneity of genetic testing availability in Australia, it will not be equitable or practical to have a concurrent control arm for comparison.” Later, they wrote, “Due to the rare nature of GKD, and because genetic testing is performed as part of routine clinical care in most states/centres, it would not be equitable to have a control arm for this study.” I was unconvinced by either of these rationales and didn’t follow their logic.

That said, the authors provide a convincing and coherent rationale in the beginning of the Discussion section (i.e., because genetic testing is clinically indicated and is standard-of-care, it would be unethical to deny genetic testing).I found the Participant Identifiers section to be a little hard to follow. Aren't most of the investigators involved with the clinical care of the patients?

Thank you for pointing this out, the main reason we have not pursued an RCT is due to the fact that genomic testing is used in many instances and is offered already in several states, and therefore it would not be ethical to deny some patients of testing for the purpose of this study. We understand that the sentences in the 'strengths and limitations section' sound a little ambiguous and have therefore amended this to better reflect our rationale highlighted in the discussion section.
Thank you, yes, most of the investigators are involved with clinical care of the patients, however some investigators are primarily involved with evaluation of the outcomes (such as project managers, health economists and researchers/students involved with survey evaluations etc). We have amended this section to make it easier to follow.

VERSION 2 – REVIEW

REVIEWER	Kurt Christensen Brigham and Women's Hospital and Harvard Medical School, USA
REVIEW RETURNED	12-Jun-2019

GENERAL COMMENTS	Please review proofs closely to clean up typos/mistakes in punctuation.