

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)	What are the factors that determine treatment choices in patients with kidney failure: a retrospective cohort study using data linkage of routinely collected data in Wales.
AUTHORS	Chess, James; Roberts, Gareth; Mc Laughlin, Leah; Williams, Gail; Noyes, Jane

VERSION 1 – REVIEW

REVIEWER	De La Mata, Nicole University of Sydney, Sydney School of Public Health
REVIEW RETURNED	18-Jul-2022

GENERAL COMMENTS	<p>Thank you for the opportunity to review this paper. The authors intended to describe and understand treatment choices following pre-dialysis education using linked data for a cohort of patients in Wales. The findings of this could be interesting and relevant to identifying opportunities to increase access to home-based dialysis therapies. However, the manuscript is poorly written which makes it cumbersome to read and grasp the main messages from the findings. There are many instances of repetition and verbose, simultaneously with lack of details on the methodological approach. The discussion is written more like a literature review of the potential barriers to accessing home-based therapy, much of which is not supported by any results presented in this study. Box 2 is a perfect example, which lists actionable questions for health professionals – none of which is presented or discussed before this point in the manuscript. I would suggest the manuscript is heavily revised to be more succinct and focused on the purpose of the paper to be easier for readers to understand the study and take-away main findings. Please note there are two different page numbers provided on each page. I am referencing the page number given on the top right corner of the document proof.</p> <p>Major comments</p> <p>Abstract, main outcome measures: Baseline demographics are not outcome measures, these are study factors. Unclear if what was discussed at first education session is also a study factor. I believe what the chosen treatment and what treatment was started are the outcome measures.</p> <p>Abstract, results. The chosen treatments do not add to 100%. What was the chosen treatment in the remaining 20%?</p> <p>Methods. This could be revised to improve the overall conciseness and flow of the methods. There are some instances of grammatical errors and repetition. Eg. Methods, setting, first sentence needs to be rephrased for grammatical correctness. Eg. “Adult patients resident in Wales...” rephrase to “Adult patients residing in Wales...”. Eg. Box 1. Please revise: “physical” to “physically”. Eg.</p>
-------------------------	--

	<p>Methods, study design. This first sentence of this paragraph is repetitive and not necessary. I suggest you delete for conciseness. Eg. Methods, setting, second sentence could be revised to “The types of education materials provided and usual processes in each of the participating centres has been reported elsewhere.” Eg. More concise if Page 5, lines 42 to 45 are combined with the following paragraph in Data sources.</p> <p>Eg. This sentence could be revised to “Our study included structure clinical data that is collected in a standardised form held in the renal electronic patient record...”. Similarly the next sentence could be revised to “We did not include the small number of patients who resident in Wales and attend NHS England renal units.”. Repetition: Methods, study design, first sentence. Methods, data sources, first sentence is similar to first sentence in previous paragraph.</p> <p>Methods, study design. Please provide details on what type of linkage was performed – deterministic or probabilistic. What personal identifiers was used to match records? I understand you have provided some references which likely provide further details on this, but a brief summary is still needed for this paper.</p> <p>Methods, page 8, lines 31-37. Could you clarify how long ago records from PEDW are? These were used to identify comorbidities, but it might not be appropriate to include comorbidities reported 10+ years since the pre-dialysis education date. Other published studies have used no more than 5 years lookback (Preen DB, Holman CDJ, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. J Clin Epidemiol 2006; 59:940-46).</p> <p>Methods, page 10, lines 5-8. This paragraph could be revised for conciseness and clarity. Eg. The primary outcomes were patient survival at 1 year and overall, and renal disease progression. Renal disease progression was considered a composite of...</p> <p>Methods, page 10, lines 5-8 (Follow-up and outcomes). You haven't clarified the follow-up period. Are all patients follow-up to death until a certain timepoint, if you are linking to death registry? Or are patients follow-up until a certain date based on their engagement with the renal health services?</p> <p>Page 10, lines 5-8. Please clarify what measures of survival are being compared. This should be described in the statistical methods section. Are these survival probabilities from Kaplan-Meier estimates or restricted mean survival times?</p> <p>Methods, page 10, lines 19-26 (Statistical methods). There is so much repetition here relating to dichotomising variables. Please revise for conciseness.</p> <p>Methods, page 10, Statistical methods. This entire section does not mention what regression method you are using at all, assuming you are using logistic regression since you have mentioned odds ratio. Why was univariate correlations used rather than univariate regression models to determine which variables to include in the multivariate regression models?</p> <p>Page 12, lines 11-16. Some of this should instead be presented in the methods (from second sentence) as it is a not a result.</p> <p>Page 12, lines 26-33. Please revise to be more succinct and direct. Eg. A higher proportion of patients were living in areas of more deprivation, with 21.8% living in the most deprived quintile compared to 16.6% living in the least deprived. Eg. Only 7.9% of patients had no additional comorbid condition, with 14.4% having 5 or more additional comorbidities.</p> <p>Page 15. The summary of results from Table 3 only discusses age. Are there any other important factors that are different to discuss? For examples, females are consistently less likely to discuss any of</p>
--	---

	<p>the treatment options, except for MCM.</p> <p>Page 16, details of discussion for patients who did not option for Peritoneal dialysis/home haemodialysis. This section provides relevant information into why PD and Home HD was not selected, but it is presented in a cumbersome block of text. In lines 31-37, you sometimes just present numbers as though it is a table without the word “in” to quote numbers. Eg “...the reasons cited for medical unsuitable 183 patients (37%), lack of carer 89 (18%)...”. A table could summarise much of this information and a short interpretation of the main finding would be more useful for readers.</p> <p>Page 17, Multivariate analysis of predictors of discussion. This paragraph does not present any estimates of odds ratio (except for comorbidity) and is very repetitive with each sentence listing each factor and stating if its “more likely” or “less likely”. I suggest authors revise to summarise the main findings and provide indication of magnitude. Eg. “Older patients were ~30% less likely to have HHD and PD discussed, and 70% less likely to have transplantation discussed, but were twice as likely to have MCM discussed.” The first two sentences should also be deleted as it repeats what was already presented in the methods.</p> <p>Table 4 and Table 5. You’ve clearly used some sort of model selection process to reduce variables in the multivariate model but have not outlined this in the methods. Was backwards or forwards selection used?</p> <p>Page 22, lines 44-60. Page 23, lines 7-18. Much of this discussion is what you perceive to be barriers to accessing home based treatment, but not necessarily supported by this current study. It also has a lot of self-citing to other studies using the SAIL data. For example, the most common reason for home HD not being discussed was being medically unsuitable (37%). Similarly, 32% of patients were PD was not discussed were thought to not be physically capable of performing dialysis. It would be more appropriate if your discussion was revised to centre around findings reported in this study. See also page 23, lines 19-32. All of which is a prelude to the limitation that a drop down list might not easily capture lived patient experiences.</p> <p>Box 2. Please explain how this study was able to support the development of questions for healthcare professionals that is presented in Box 2. This does not at all seem related to the results presented.</p> <p>Page 25, lines 8-13. The purpose of this study was not to link together the whole system study undertaken to understand patient experiences, shared decision making, economics and sustainability of kidney services in Wales. If this is the objective, then a different form of publication (such as literature review, opinion piece, letter to editor) would be more suitable. The purpose of this study was to use linked data to describe and understand treatment choices following pre-dialysis education.</p> <p>Minor comments</p> <p>Throughout you interchange between “kidney” and “renal”. The Kidney Disease: Improving Global Outcomes (KDIGO) recommend using the term “kidney” rather than “renal” (see https://www.kidney-international.org/article/S0085-2538%2820%2930233-7/fulltext).</p> <p>Throughout the manuscript you also interchange between “subjects”, “participants”, “patients”. I suggest you choose one and use it consistently.</p> <p>Page 8, line 23-24. Please spell out the full word first then acronym in brackets for MDRD, PEDW and CFS. Similarly, introduce the acronym eGFR at the first instance of the word estimated glomerular</p>
--	--

	<p>filtration rate. Note you also need to add the word 'rate'. Also, WRCN given on Page 10, lines 54-55.</p> <p>Page 8, lines 55-60. Would reporting the kilometres to the nearest tertiary renal centre also be useful?</p> <p>Page 10, lines 19-20. For ease of clinical interpretation, it would be more appropriate to categorise BMI into the four commonly used groups: underweight, normal, overweight and obese. Rather than use BMI as a continuous measure in 5 unit intervals.</p> <p>Page 10, lines 32-33. "We used likelihood ratios to evaluate models of multiple regression." Please clarify. Likelihood ratio tests can be used to compare models to determine if a variable is significantly associated with the outcome, as opposed to Wald tests.</p> <p>Page 12, first sentence is repetitive of what has already been covered in the methods.</p> <p>Page 12, lines 8-9. The median and interquartile range (or minimum and maximum) follow-up time would be more useful to understand distribution of follow-up for participants.</p> <p>Page 14, lines 5-7. Please revise first sentence for conciseness. Is this sentence even needed, given it is repeated in the title for table 2.</p> <p>Page 14, lines 9-11. Please revise last two sentences of paragraph to be more succinct. Eg. 23% of patients were undecided at first pre-dialysis education session, of which 17% had an additional dialysis education session.</p> <p>Page 14, lines 14-16. Please revise to be grammatical correct: "...than patients who patients who chose home dialysis or transplantation."</p> <p>Table 2. Values that were meant to be in the brackets have not been presented. Eg. (SD) for age in years, (%) for female gender. Median values should be accompanied with interquartile range (IQR).</p> <p>Page 14, lines 17-20. Please revise sentences to be more succinct and not repetitive.</p> <p>Page 16, lines 18-25. This sentence is far too long and should be broken into smaller sentences.</p> <p>Page 16, lines 31-32. Please revise to avoid repetition of "record".</p> <p>Page 18, first paragraph is repetitive of what has already been presented in Table 2 and discussed previously.</p> <p>Page 18, lines 10-13. Repetitive of what has already been discussed in methods.</p> <p>Page 20, lines 32-35. By definition, there would be a higher rate of progression in all groups other than MCM. What is the main finding of this?</p>
--	--

REVIEWER	Sood, Manish Ottawa Hospital Research Institute, Nephrology
REVIEW RETURNED	05-Oct-2022

GENERAL COMMENTS	<p>The investigators conducted a population-based retrospective cohort study examining factors that determined dialysis modality choices following pre-dialysis education in Wales, UK. Examining 1207 patients, they found 25% choose home dialysis, 32% hospital-based and 23% were undecided. Lower home dialysis was associated with age and frailty and among those who choose home dialysis, almost half did not start on it. They conclude improvement in pre-dialysis education.</p> <p>Topic of high interest and intrinsically difficult to study and in that regard, I congratulate the investigators on their work. Strengths of the study include it was population-based, a relatively large cohort, full follow up on the final dialysis modality chosen and a well written</p>
-------------------------	--

	<p>manuscript.</p> <p>Suggestions/clarifications/revisions:</p> <ul style="list-style-type: none"> - abstract conclusions are speculative and I would simply summarize the key findings - please include some details on the pre-dialysis education: was it standardized? how many centres were involved? was the instruction in English only? How many sessions? I understand the authors referenced an excellent previous study (strength) on the pre-dialysis education process and how it was developed - add 2-3 sentences explaining it to the reader without having to look it up - did you account for death or patients who ultimately choose conservative therapy? - I think not examining repeat education sessions (limiting to only the first) is a missed opportunity and the investigators should examine the impact of repeat sessions - not sure why the MDRD education was used when the CKD-epi is more accurate and considered the standard; ideally should use the updated race-free version - why did the authors not exclude (or better yet, stratify by) individuals on KRT at the time of education? I assume these are mostly in centre HD patients; this is a very different population including AKI - in the analysis the dichotomization of many continuous variables resulted in the loss of valuable information and non-linear associations; they should examine them ideally with non functional forms or at least categorically - the degree of missing data needs to be clearly presented; complete case analysis may not be suitable depending on the pattern of missingness and other techniques such as imputations may be required - no details on when or who was censored for survival analysis - inclusion of variables in the multivariate model based on univariate statistical significance is not the ideal means of identifying factors; consider variable selection based on clinical importance and or previous literature - in general the analysis section is too brief and lacks detail for reproducibility - not all patients are candidates for home dialysis due to cognition or physical limitations, absolute contraindications in PD - where they excluded?
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
 Dr. Nicole De La Mata, University of Sydney Comments to the Author:
 Thank you for the opportunity to review this paper. The authors intended to describe and understand treatment choices following pre-dialysis education using linked data for a cohort of patients in Wales. The findings of this could be interesting and relevant to identifying opportunities to increase access to home-based dialysis therapies. However, the manuscript is poorly written which makes it cumbersome to read and grasp the main messages from the findings. There are many instances of repetition and verbose, simultaneously with lack of details on the methodological approach. The discussion is written more like a literature review of the potential barriers to accessing home-based therapy, much of which is not supported by any results presented in this study. Box 2 is a perfect example, which lists actionable questions for health professionals – none

of which is presented or discussed before this point in the manuscript. I would suggest the manuscript is heavily revised to be more succinct and focused on the purpose of the paper to be easier for readers to understand the study and take-away main findings.

Response: Thank you for reviewing and the feedback. We have removed box 2 and revised text to remove repetition.

Major comments

Abstract, main outcome measures: Baseline demographics are not outcome measures, these are study factors. Unclear if what was discussed at first education session is also a study factor. I believe what the chosen treatment and what treatment was started are the outcome measures. Abstract, results. The chosen treatments do not add to 100%. What was the chosen treatment in the remaining 20%?

Response: Thank you, we have revised the outcome measures. We have also revised the treatments, which were omitted for brevity, so they add up to 100%

Methods. This could be revised to improve the overall conciseness and flow of the methods. There are some instances of grammatical errors and repetition. Eg. Methods, setting, first sentence needs to be rephrased for grammatical correctness. Eg. "Adult patients resident in Wales..." rephrase to "Adult patients residing in Wales...". Eg. Box 1. Please revise: "physical" to "physically". Eg. Methods, study design. This first sentence of this paragraph is repetitive and not necessary. I suggest you delete for conciseness. Eg. Methods, setting, second sentence could be revised to "The types of education materials provided and usual processes in each of the participating centres has been reported elsewhere." Eg. More concise if Page 5, lines 42 to 45 are combined with the following paragraph in Data sources.

Eg. This sentence could be revised to "Our study included structure clinical data that is collected in a standardised form held in the renal electronic patient record...". Similarly the next sentence could be revised to "We did not include the small number of patients who resident in Wales and attend NHS England renal units.". Repetition: Methods, study design, first sentence. Methods, data sources, first sentence is similar to first sentence in previous paragraph.

Response: We have made the changes suggested and standardised terms.

Methods, study design. Please provide details on what type of linkage was performed – deterministic or probabilistic. What personal identifiers was used to match records? I understand you have provided some references which likely provide further details on this, but a brief summary is still needed for this paper.

Response: SAIL uses a combination of deterministic and probabilistic matching. Records are first matched on a deterministic basis using NHS number followed by name, surname, date of birth, sex and finally postcode. Any unmatched records are subject to probabilistic matching. We have updated the text to reflect this.

Methods, page 8, lines 31-37. Could you clarify how long ago records from PEDW are? These were used to identify comorbidities, but it might not be appropriate to include comorbidities reported 10+ years since the pre-dialysis education date. Other published studies have used no more than 5

years lookback (Preen DB, Holman CDJ, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. J Clin Epidemiol 2006; 59:940-46).

Response: Thank you for the comment. Our method utilised did not limit the lookback period for comorbidities which were derived both from hospital admissions and the primary care record. We acknowledge the provided reference describes superior performance at predicting post-hospital mortality when using a 1 year lookback period. We have clarified the lookback period in the methods text.

Methods, page 10, lines 5-8. This paragraph could be revised for conciseness and clarity. Eg. The primary outcomes were patient survival at 1 year and overall, and renal disease progression. Renal disease progression was considered a composite of...

Response: We have revised the methods paragraph as suggested.

Methods, page 10, lines 5-8 (Follow-up and outcomes). You haven't clarified the follow-up period. Are all patients follow-up to death until a certain timepoint, if you are linking to death registry? Or are patients follow-up until a certain date based on their engagement with the renal health services?

Response: The duration of follow up was determined by the coverage of the linked data. We used source electronic patient record and WDS (our national demographic registry) to confirm date of death. We followed all patients up to a specific date, which provided us with a minimum follow up period of 586 days after pre-dialysis education. We have added this paragraph to "Follow up and Outcomes" to help clarify: -

"We followed patients up until a set date determined by coverage by linked data, this resulted in a minimum follow up of 586 days for each patient from the date of their first pre-dialysis education session."

To prevent repetition we removed this sentence from the first results paragraph: -

"The minimum observed follow up for a participant was 586 days."

Page 10, lines 5-8. Please clarify what measures of survival are being compared. This should be described in the statistical methods section. Are these survival probabilities from Kaplan-Meier estimates or restricted mean survival times?

Response: We described simple patient survival at one year, as well as overall survival at time of analysis of study. We did not model survival or correct for known predictors of survival such as age or comorbidity. We intended to describe the simple 1 year survival in a descriptive manner.

We have added a sentence in the statistical methods section:- "Patient survival was calculated at

one year following pre-dialysis education as the proportion of patients still alive. No correction was made for age or comorbidity.”

Methods, page 10, lines 19-26 (Statistical methods). There is so much repetition here relating to dichotomising variables. Please revise for conciseness.

Response: We have revised the section.

Methods, page 10, Statistical methods. This entire section does not mention what regression method you are using at all, assuming you are using logistic regression since you have mentioned odds ratio

Response: Thank you, we have added this to statistical methods.

Why was univariate correlations used rather than univariate regression models to determine which variables to include in the multivariate regression models?

Response: We used univariate regression, we have corrected the text.

Page 12, lines 11-16. Some of this should instead be presented in the methods (from second sentence) as it is a not a result.

Response: We have moved this paragraph to methods.

Page 12, lines 26-33. Please revise to be more succinct and direct. Eg. A higher proportion of patients were living in areas of more deprivation, with 21.8% living in the most deprived quintile compared to 16.6% living in the least deprived. Eg. Only 7.9% of patients had no additional comorbid condition, with 14.4% having 5 or more additional comorbidities.

Response: Thank you, we have made suggested changes.

Page 15. The summary of results from Table 3 only discusses age. Are there any other important factors that are different to discuss? For examples, females are consistently less likely to discuss any of the treatment options, except for MCM.

Response: There was a typing error in that table which has been corrected. The proportions of patients with female gender were similar across groups and representative of the cohort overall. We have added a paragraph to describe likelihood that the options were discussed/not-discussed: -

The likelihood that the various options were discussed did not appear to be influenced by patient gender, travelling time to the unit, deprivation, BMI or eGFR. There appeared to be more chance that MCM was discussed, or home dialysis options not discussed where the EQ-5D or VAS score

was lower or there were more comorbidities or inpatient hospitalisation days.

Page 16, details of discussion for patients who did not option for Peritoneal dialysis/home haemodialysis. This section provides relevant information into why PD and Home HD was not selected, but it is presented in a cumbersome block of text. In lines 31-37, you sometimes just present numbers as though it is a table without the word “in” to quote numbers. Eg “...the reasons cited for medical unsuitable 183 patients (37%), lack of carer 89 (18%)...”. A table could summarise much of this information and a short interpretation of the main finding would be more useful for readers.

Response: Thank you for the feedback. We have created a new table and changed the discussion to reflect the table.

Page 17, Multivariate analysis of predictors of discussion. This paragraph does not present any estimates of odds ratio (except for comorbidity) and is very repetitive with each sentence listing each factor and stating if its “more likely” or “less likely”. I suggest authors revise to summarise the main findings and provide indication of magnitude. Eg. “Older patients were ~30% less likely to have HHD and PD discussed, and 70% less likely to have transplantation discussed, but were twice as likely to have MCM discussed.” The first two sentences should also be deleted as it repeats what was already presented in the methods.

Response: Thank you for the feedback, we have made the changes suggested.

Table 4 and Table 5. You've clearly used some sort of model selection process to reduce variables in the multivariate model but have not outlined this in the methods. Was backwards or forwards selection used?

Response: We used forward stepwise method using likelihood ratios. We have revised the methods to reflect this.

Page 22, lines 44-60. Page 23, lines 7-18. Much of this discussion is what you perceive to be barriers to accessing home based treatment, but not necessarily supported by this current study. It also has a lot of self-citing to other studies using the SAIL data. For example, the most common reason for home HD not being discussed was being medically unsuitable (37%). Similarly, 32% of patients were PD was not discussed were thought to not be physically capable of performing dialysis. It would be more appropriate if your discussion was revised to centre around findings reported in this study. See also page 23, lines 19-32. All of which is a prelude to the limitation that a drop down list might not easily capture lived patient experiences.

Response: Thank you we have condensed and revised this section.

Box 2. Please explain how this study was able to support the development of questions for healthcare professionals that is presented in Box 2. This does not at all seem related to the results presented.

Response: We have removed box 2

Page 25, lines 8-13. The purpose of this study was not to link together the whole system study undertaken to understand patient experiences, shared decision making, economics and sustainability of kidney services in Wales. If this is the objective, then a different form of publication (such as literature review, opinion piece, letter to editor) would be more suitable. The purpose of this study was to use linked data to describe and understand treatment choices following pre-dialysis education.

Response: Thank you. We have comprehensively reviewed and reorganised the manuscript in line with BMJ Open template and reorientated the focus as per your suggestion.

Minor comments

Throughout you interchange between “kidney” and “renal”. The Kidney Disease: Improving Global Outcomes (KDIGO) recommend using the term “kidney” rather than “renal” (see) Throughout the manuscript you also interchange between “subjects”, “participants”, “patients”. I suggest you choose one and use it consistently.

Response: Thank you for the feedback. We have chosen consistent terms.

Page 8, line 23-24. Please spell out the full word first then acronym in brackets for MDRD, PEDW and CFS. Similarly, introduce the acronym eGFR at the first instance of the word estimated glomerular filtration rate. Note you also need to add the word ‘rate’. Also, WRCN given on Page 10, lines 54-55.

Response: Thank you, we have made suggested corrections.

Page 8, lines 55-60. Would reporting the kilometres to the nearest tertiary renal centre also be useful?

Response: We did calculate this metric within the data safe haven, which had similar findings. In the interest of brevity we have not prevented the data.

Page 10, lines 19-20. For ease of clinical interpretation, it would be more appropriate to categorise BMI into the four commonly used groups: underweight, normal, overweight and obese. Rather than use BMI as a continuous measure in 5 unit intervals.

Response: As a team we debated this, and whilst it would be easier to interpret it would have complicated the regression techniques used, requiring the use of ordinal technique and lead to loss of granularity of data by ascertaining patients to groups for a continuous variable.

Page 10, lines 32-33. “We used likelihood ratios to evaluate models of multiple regression.”. Please clarify. Likelihood ratio tests can be used to compare models to determine if a variable is

significantly associated with the outcome, as opposed to Wald tests.

Response: We used the inbuilt model selection criteria built into SPSS, using the forward stepwise technique which used likelihood ratio.

Page 12, first sentence is repetitive of what has already been covered in the methods.

Response: We have removed the repetitive sentence.

Page 12, lines 8-9. The median and interquartile range (or minimum and maximum) follow-up time would be more useful to understand distribution of follow-up for participants.

Response: Thank you for the suggestion, we have changed how we describe follow-up.

Page 14, lines 5-7. Please revise first sentence for conciseness. Is this sentence even needed, given it is repeated in the title for table 2.

Response: Thank you for the suggestion, we have made the suggested changes.

Page 14, lines 9-11. Please revise last two sentences of paragraph to be more succinct. Eg. 23% of patients were undecided at first pre-dialysis education session, of which 17% had an additional dialysis education session.

Response: We have accepted the suggested revision.

Page 14, lines 14-16. Please revise to be grammatical correct: "...than patients who patients who chose home dialysis or transplantation."

Response: We have accepted the suggested revision.

Table 2. Values that were meant to be in the brackets have not been presented. Eg. (SD) for age in years, (%) for female gender. Median values should be accompanied with interquartile range (IQR).

Response: We have moved the references to the values in brackets

Page 14, lines 17-20. Please revise sentences to be more succinct and not repetitive.

Response: We have revised the sentences.

Page 16, lines 18-25. This sentence is far too long and should be broken into smaller sentences.

Response: We have revised this section and introduced a table.

Page 16, lines 31-32. Please revise to avoid repetition of "record".

Response: We have revised this section.

Page 18, first paragraph is repetitive of what has already been presented in Table 2 and discussed previously.

Response: We have removed the repetitive sentence.

Page 18, lines 10-13. Repetitive of what has already been discussed in methods.

Response: We have removed the repetitive section.

Page 20, lines 32-35. By definition, there would be a higher rate of progression in all groups other than MCM. What is the main finding of this?

Response: Presumably patients who chose MCM passed away with comorbid disease before progressing to reach an eGFR <8. We have added a sentence in the discussion to address this.

Reviewer: 2

Dr. Manish Sood, Ottawa Hospital Research Institute Comments to the Author:

The investigators conducted a population-based retrospective cohort study examining factors that determined dialysis modality choices following pre-dialysis education in Wales, UK. Examining 1207 patients, they found 25% choose home dialysis, 32% hospital-based and 23% were undecided. Lower home dialysis was associated with age and frailty and among those who choose home dialysis, almost half did not start on it. They conclude improvement in pre-dialysis education.

Topic of high interest and intrinsically difficult to study and in that regard, I congratulate the investigators on their work. Strengths of the study include it was population-based, a relatively large cohort, full follow up on the final dialysis modality chosen and a well written manuscript.

Suggestions/clarifications/revisions:

- abstract conclusions are speculative and I would simply summarize the key findings

Response: Thank you reviewing our paper. We have removed the last sentence to remove the speculative findings.

- please include some details on the pre-dialysis education: was it standardized? how many centres were involved? was the instruction in English only? How many sessions? I understand the authors referenced an excellent previous study (strength) on the pre-dialysis education process and how it was developed - add 2-3 sentences explaining it to the reader without having to look it up

Response: Thank you very much for this important observation. We did not record the language used for the interview but is likely to be in either English or Welsh. We have published this work in detail elsewhere. In this paper we have added a sentence in the introduction explaining the ideal SDM model and highlighted in the discussion, that implementation has been variable and it not consistently delivering the desired outcomes. We have limited additional details to ensure this papers core focus and agenda is clear as per all reviewers comments and suggestions.

- did you account for death or patients who ultimately choose conservative therapy?

Response: We did record patients who initially chose conservative therapy, we have described how they have lower 1 year survival(62.3%) and during the follow up period 72.7% passed away. A low rate of patients who chose MCM/conservative care had progressive CKD (eg reached eGFR <8) presumably as they passed away due to comorbid disease. Few patients who chose MCM went onto have KRT.

- I think not examining repeat education sessions (limiting to only the first) is a missed opportunity and the investigators should examine the impact of repeat sessions

Response: We agree with the reviewer that we should examine the impact of repeat sessions. We currently feel the manuscript has a large amount of data to present to the reader and felt additional information may overload the reader. We wanted to prioritise publishing the data on the first session before analysing subsequent sessions. As authors we feel that multiple education sessions, where resources permit, may help patients who are initially undecided and/or unsure chose a home dialysis option.

not sure why the MDRD education was used when the CKD-epi is more accurate and considered the standard; ideally should use the updated race-free version

Response: We used the race-free MDRD equation as this was in-use at the time of the study in clinical practice.

why did the authors not exclude (or better yet, stratify by) individuals on KRT at the time of education? I assume these are mostly in centre HD patients; this is a very different population including AKI

Response: Usual practice is to offer patients dialysis education with progressive CKD, or patients where AKI appears to be sustained and recovery is not expected. I can confirm this group of patients were in centre HD patients. We appreciate this cohort of patients on KRT is a different population and agree there would be merits to analysing this group separately. In the interests of keeping the analysis manageable we did not analyse this population separately.

- in the analysis the dichotomization of many continuous variables resulted in the loss of valuable information and non-linear associations; they should examine them ideally with non functional forms or at least categorically

Response: We tried to avoid the loss of granularity in data that comes with categorisation of variables. We did however feel we needed to do this to both aid clarity and interpretation (For instance in the baseline demographics table), but mostly to help the statistical analysis. To keep the complexity of the statistical analysis manageable dichotomisation to create a dummy binary variable allowed us to analyse the data using logistic regression and avoid using less flexible ordinal regression techniques.

the degree of missing data needs to be clearly presented; complete case analysis may not be suitable depending on the pattern of missingness and other techniques such as imputations may be required

Response: We feel we have documented the degree of missing data in each table. In table one we present the number of the original 1707 cohort and record where we have data available. For example when it came to comorbidity data we had data for 1643/1707 patients eg 96%.

When analysing the data on a multivariate basis we did undertake a complete case analysis.

We did not find a particular pattern of missingness and so have assumed the data is MAR (missing at random). In the interest of keeping the analysis as simple as possible we have avoided imputation or modelling the missingness explicitly.

- no details on when or who was censored for survival analysis

Response: The focus of our work has not been to explicitly model survival between groups, but we felt it important to convey a measure to describe survival between groups of patients on the basis of the option they chose. We picked 1 year survival as we knew we had no loss of follow up in the first year and good quality linkage of data for the event of death. We did not censor any patients either at the time of renal transplant or starting kidney replacement therapy.

- inclusion of variables in the multivariate model based on univariate statistical significance is not

the ideal means of identifying factors; consider variable selection based on clinical importance and or previous literature
Response: We agree the inclusion of variables in the model is not always thought to be the optimal technique. We did explore the data and on occasions forced clinically important variables into the regression model. On the occasions we did this it did not result in a parsimonious model

- in general the analysis section is too brief and lacks detail for reproducibility
Response: We feel we have been limited by the quantity of data and tables we wished to include along with the word count. As our data is pseudo-anonymised and held in data safe haven with strict rules limiting which data can be exported, we would not be able to open source the data or analysis to allow the analysis to be reproduced outside of the SAIL research environment.

- not all patients are candidates for home dialysis due to cognition or physical limitations, absolute contraindications in PD - where they excluded?
Response: We did not specifically exclude patients from this study due to these limitations, though it is possible that there were patients with CKD & significant cognition or physical limitations that were never referred to the renal team or for pre dialysis education as it would never be practicable. We have added an additional table to make it clear the proportion of patients where home dialysis was not discussed with patients due to these limitations, or where a patient did not choose a home therapy for similar reasons. (Table 4)

VERSION 2 – REVIEW

REVIEWER	Sood, Manish Ottawa Hospital Research Institute, Nephrology
REVIEW RETURNED	19-Dec-2023
GENERAL COMMENTS	reviewers comments were addressed