PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Effect of an integrated care pathway on use of primary and secondary health care by patients at high-risk of emergency inpatient admission: a matched control cohort study in Tower Hamlets
AUTHORS	Parry, Will; Wolters, Arne; Brine, Richard; Steventon, Adam

VERSION 1 - REVIEW

REVIEWER	Suzanne Mason ScHARR, University of Sheffield, UK
REVIEW RETURNED	19-Sep-2018

GENERAL COMMENTS	This is an exceedingly long (over word limit?) and possibly overly complex paper on the impact of integrated care on subsequent healthcare utilisation. I have the following comments in no particular order: 1. I think the authors should consider splitting this into two papers - there is a strong emphasis on the methods applied to match the two groups of patients. I am afraid this seems to overshadows the analysis and dilutes the clinical / service delivery message. I think if there were a methods paper that can be referenced in this paper it would reduce the length of the paper and the complexity and make it far more accessible to clinicians and service providers. 2. The single site analysis in London limits any form of generalisability for the findings. 3. The findings may lack statistical significance, but the authors should consider clinical and service significance attached to some of the observed changes in the outcomes measured. Just because it lacks statistical significance does not mean that a change is not worth having! 4. The lack of statistical significance could be because there was no power calculation undertaken to decide how many patients would be required in each group. I suspect the study is underpowered for the primary outcome. 5. I noted there was no ethical approval sought for this study - I think a statement explaining why would be helpful. 6. There are too many charts - I think some of them could be dispensed with or added to a web appendix.

REVIEWER	Rachael Hunter University College London, UK
REVIEW RETURNED	24-Oct-2018

GENERAL COMMENTS	This is an incredibly interesting analysis of an integrated care pathway in Tower Hamlets. The authors have put a lot of work into trying to ensure bias due to the observational nature of the study is minimised. It's a shame they were unable to locate another area as a control as this would have made this a far stronger study. It's clear though from the paper the challenges they have faced and that they have done the best they can to overcome this. This does not change the fact that there is the potential for the patients who accessed integrated care to be systematically different to those who did not in an unobservable way. The authors have sufficiently noted this limitation in their discussion though. I have the following comments to the authors: 1) An important component of the results is that the increased attendances for elective inpatient admissions may have been because of unmet need being identified, which the authors have nicely pointed out on page 23. As a result collecting long term outcomes for a study like this is important, and potentially the 1 year follow-up and outcomes chosen were insufficient to adequately evaluate the important mortality and morbidity outcomes of the service. That the duration of the study was potentially insufficient to identify important clinical benefits maybe should be noted more strongly in the abstract and conclusion. 2) The paper would benefit from a clear aim of the work at the end of the introduction. 3) The paper would also benefit from more clearly stating the results of the analysis in relation to the aim in the first sentence of the discussion. 4) A lot of the last paragraph in the methods section "Integrated care" (page 6 line 44 onwards) gives a description of integrated care that would be better suited to being placed in the introduction.

REVIEWER	Harriet Hiscock Murdoch Children's Research Institute Australia
REVIEW RETURNED	28-Oct-2018

GENERAL COMMENTS	I enjoyed reading this paper. I have some concerns that need to be addressed: 1. Abstract: please state the outcome for your primary endpoint first in the Results. 2. Methods: can you please include a sample size calculation, as per your protocol. I note that you did not achieve your stated sample size - this should be included in your Discussion as a limitation with an acknowledgement that you may have been underpowered to detect differences in your outcomes between cases and controls. In your methodology, there is likely to be an effect of clustering at the level of the individual GP as well as the level of the GP practice, on outcomes. Have you accounted for clustering (I could not see this)? If not, why not? How many hospitals were included in your analyses? For international readers, can you clarify that you will have captured all

	relevant hospitals for your patient group? Will patients have gone to hospitals outside the study area, as they may well do in Australia? 3. Discussion: there are several other limitations that need to be addressed, including: - did you take into account duration of IC that the cases received (as per your protocol)? Could it be that there was no effect seen on admissions because of insufficient IC dose and duration? -please also mention your power. - how did the 30% of IC cases with non-matched controls differ from the 70% with matched controls on baseline variables? Was there any evidence of differences in potential confounding variables? Can you also include in the Discussion your: - future research recommendations (eg longer follow up, larger sample, subgroup analyses of patients with/without a mental health condition, inclusion of other outcomes including re-admission rates and costs), as proposed in your protocol? - policy recommendations - what should commissioners do now, based on your findings?

REVIEWER	Ian Duncan Dept. of Statistics & Applied Probability University of California Santa Barbara California U.S.A.
REVIEW RETURNED	23-Dec-2018

GENERAL COMMENTS	I like this paper a lot for two reasons: 1. The authors are willing to publish a paper without positive outcomes. Given publication bias, this happens rarely, particularly with this type of intervention, and so makes a considerable contribution to the literature. 2. The statistical analysis is unusually robust and described in some detail. Indeed the analytical portion of the paper alone makes this paper worthwhile reading. If I have a quibble it is the length of the paper. It could well be a chapter in a book. The introduction, for example, runs 2-1/2 pages. The introduction includes an interesting history of care management efforts. I will let the editor decide whether this is appropriate for the journal. I assume that the attachment – the protocol "Evaluation of the Integrated Care Program ….Waltham Forest et al."- will not form part of the actual publication? Some questions arose as I reviewed the paper. 1. The endpoints of the study make sense and those that I would expect (A&E, bed days etc.). I am a bit surprised by the increase in elective admissions and would appreciate more detail, such as: what admissions are considered elective vs. non-elective? In the U.S. we consider elective admissions to be those for elective procedures, i.e. non-emergency, and sometimes scheduled admissions (e.g. for repair of a hip replacement). An increase in truly elective admissions suggests some element of induced demand, which is the reverse of what one would expect with this type of management program. Perhaps the increased contact with the medical system is responsible for inducing demand, an hypothesis that the authors raise toward the end of the paper. 2. I have not studied the QAdmissions risk score paper recently, but will try to find time to do so. In thinking about what could be driving the counter-intuitive results, I wonder whether this has to

	do with the risk score methodology? In the U.S. we observe fairly significant regression to the mean in terms of risk scores in elderly populations. One reason for this is that the risk score is diagnosis driven, and hospital records contain much more detailed diagnoses, so a patient who has had a recent admission is more likely to have a higher risk score (and not to have one the following year if not re-admitted). Regression does not seem to be the issue with this study, since elective admissions actually increased, but I am left wondering about the calculation and role of the risk score. 3. The authors have used Kaplan-Meier analysis to check the survival of the intervention and control groups, which is appropriate. Survival, however, raises a different question: do the authors have data to explore survival longitudinally? Continuing to puzzle over the results and the higher use of elective admissions, I wonder whether this in any way contributes to longer survival of the intervention population? The authors would likely need a few years of longitudinal data to assess survival. Thinking about analysis of longitudinal data I wonder also what the pattern of risk scores looks like, post-program? I am not suggesting that the authors add to the analysis, but this could be a topic for a subsequent paper.

REVIEWER	Anup Karan Indian Institute of Public Health-Delhi (IIPHD), Public Health Foundation of India (PHFI)
REVIEW RETURNED	30-Dec-2018

GENERAL COMMENTS	Effect of an integrated care pathway on use of primary and secondary health care by patients at high-risk of emergency inpatient admission: a matched control cohort study in Tower Hamlets Comments: This study estimates effect of an integrated care pathway on primary and secondary healthcare use by high-risk of emergency inpatient in Tower Hamlets, a deprived, inner city London borough. The study uses real-time person-level administrative data and estimates effects of a range of interventions, such as case management; support with self-care; discharge support; enhanced care coordination; and specialist input in community settings on endpoints namely numbers of emergency inpatient admissions and elective inpatient admissions, inpatient bed days, accident and emergency attendances, outpatient attendances and general practitioner contacts in the year after enrolment. The study uses improved techniques of matching and create comparable sample before applying mean difference tests and difference-in-differences regression analysis. The methods also claim to control roles of unobserved covariates. On the whole, the methods in the study looks seemingly robust leading to unbiased results. I don't have any major comments on the paper. However, I have a few minor observations which authors may like to address. May main comments are as follows: Minor Comments:

	1. After complete matching the number of observation left for final analysis are 1,720 in treatment group and 2,867 in matched control group. This reflects a large reduction in the original sample of 10,411 in the treatment and 29,885. Even after making all ineligible participants the treatment and control samples were 2,459 and 97,040 respectively before matching. Matching left very small sample. I just to flag that this might create some kind of skewed sample and may have implications on the generalibility of the results. I would like to know reply of authors on this issue else this can be mentioned in the limitation of the study. 2. The entire method section of the paper, although looks robust, looks little crowded. I would like to send some materials from the method section into supplementary material. For instance, the matching algorithm presented in Box 1 may be shifted to supplementary material. Authors may also like to condense the method section by putting some other details in supplementary materials. 3. All the figures do not carry the Figure numbers and the required titles as to make one-to-one correspondence with the text in the manuscript. 4. I think the main results of the study are in Table 3 and Table 4. I guess, putting Table 1 and Table 2 and the related descriptions in the supplementary file will improve the readability of the paper. The main manuscript may include only brief results of the matching. 5. The section on 'statement of findings' just consolidate the main findings. I would suggest authors to provide some probable explanations of almost no effects of the interventions on the endpoints. Also, can authors provide their opinion if such interventions are expected to affects the endpoints in general or not?

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Suzanne Mason

Institution and Country: ScHARR, University of Sheffield, UK

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

Please leave your comments for the authors below

This is an exceedingly long (over word limit?) and possibly overly complex paper on the impact of integrated care on subsequent healthcare utilisation.

I have the following comments in no particular order:

1. I think the authors should consider splitting this into two papers - there is a strong emphasis on the methods applied to match the two groups of patients. I am afraid this seems to overshadows the analysis and dilutes the clinical / service delivery message. I think if there were a methods paper that can be referenced in this paper it would reduce the length of the paper and the complexity and make it far more accessible to clinicians and service providers.

Although we make use of novel methods, the paper does not include any completely new methodology, and so we do not feel a separate methods paper would be appropriate in this case. However, we acknowledge that the paper is on the long side, as other reviewers have also commented. We thought this would be acceptable for readers, as the BMJ Open is an easily navigable web journal, and so sections of primary interest can be identified easily. However, we have decided to move some of the methods detail (on the matching algorithm) to the supplementary material, and reference it in the main text. We hope this is sufficient.

2. The single site analysis in London limits any form of generalisability for the findings.

Yes, we agree with the reviewer on this point – the findings are specific to a deprived London borough. We have highlighted this limitation.

3. The findings may lack statistical significance, but the authors should consider clinical and service significance attached to some of the observed changes in the outcomes measured. Just because it lacks statistical significance does not mean that a change is not worth having!

The aim of the analysis was to identify effects on endpoints which could be attributed to the use of integrated care in Tower Hamlets. Where findings have very wide confidence intervals associated with them, we do not feel there is sufficient evidence to suggest that the intervention had a beneficial effect on these endpoints by the end of the study period, considering that it is a quasi-experimental study. However, we have included reference to the power of the study as an issue.

4. The lack of statistical significance could be because there was no power calculation undertaken to decide how many patients would be required in each group. I suspect the study is underpowered for the primary outcome.

The study was somewhat underpowered compared to the power calculation undertaken in the original protocol (see supplementary material), as two out of three sites were not included in the analysis, and many control patients were of too-low risk to include. We have highlighted this limitation and included revised power calculations in the supplementary material, however we do not generally think it wise to put too much focus on post-hoc power calculations, as the achieved power is demonstrated by the widths of the resulting confidence intervals for the endpoints in any case. The power values for the models used in this study are related to the sample sizes for the two groups, the variance and overdispersion of the endpoints, the duration of the study, and the type of model and the association with covariates employed as controls. The effect of the non-sample size elements of power can be seen in the difference between the confidence intervals for GP contacts and inpatient endpoints, for example.

5. I noted there was no ethical approval sought for this study - I think a statement explaining why would be helpful.

Ethical approval was not sought as the study involved retrospective analysis of existing pseudonymised administrative data, for the purposes of service evaluation. However, the original protocol was approved by the Waltham Forest and East London Collaborative steering group on 16th June 2015 (see original protocol in supplementary material). We have added a statement to this effect in the text.

6. There are too many charts - I think some of them could be dispensed with or added to a web appendix.

We feel that this is not so relevant to online web journals like the BMJ Open as it is to print journals. We note that BMJ Open routinely present Figures and Tables in web GUI ‘accordion’ elements, thus readers can choose to view each figure and table if relevant to their interests, or ignore them and continue with the text.

Reviewer: 2

Reviewer Name: Rachael Hunter

Institution and Country: University College London, UK

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

Please leave your comments for the authors below

This is an incredibly interesting analysis of an integrated care pathway in Tower Hamlets. The authors have put a lot of work into trying to ensure bias due to the observational nature of the study is minimised. It's a shame they were unable to locate another area as a control as this would have made this a far stronger study. It's clear though from the paper the challenges they have faced and that they have done the best they can to overcome this. This does not change the fact that there is the potential for the patients who accessed integrated care to be systematically different to those who did not in an unobservable way. The authors have sufficiently noted this limitation in their discussion though.

Thank you. We also feel we have made the most of a difficult situation, considering the reduced scope compared to the original protocol (one site instead of three), and the problems finding sufficient control patients with similar risk to the integrated care patients.

I have the following comments to the authors:

1) An important component of the results is that the increased attendances for elective inpatient admissions may have been because of unmet need being identified, which the authors have nicely pointed out on page 23. As a result collecting long term outcomes for a study like this is important, and potentially the 1 year follow-up and outcomes chosen were insufficient to adequately evaluate the important mortality and morbidity outcomes of the service. That the duration of the study was potentially insufficient to identify important clinical benefits maybe should be noted more strongly in the abstract and conclusion.

Yes, we agree that this is a substantial limitation. We have emphasised in the strengths and limitations and conclusion that a longer follow-up was not possible, and that this limited any potential to identify improvements to endpoints and longevity of patients over the medium and long-term.

2) The paper would benefit from a clear aim of the work at the end of the introduction.

Agreed. We have added a clear aim to the introduction.

3) The paper would also benefit from more clearly stating the results of the analysis in relation to the aim in the first sentence of the discussion.

Agreed. We have added this to the start of the discussion.

4) A lot of the last paragraph in the methods section "Integrated care" (page 6 line 44 onwards) gives a description of integrated care that would be better suited to being placed in the introduction.

Agreed. We have moved some text from the Methods to the Introduction.

Reviewer: 3

Reviewer Name: Harriet Hiscock

Institution and Country: Murdoch Children's Research Institute - Australia

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

Please leave your comments for the authors below

I enjoyed reading this paper. I have some concerns that need to be addressed:

1. Abstract: please state the outcome for your primary endpoint first in the Results.

Agreed. We have amended the results section.

2. Methods: can you please include a sample size calculation, as per your protocol. I note that you did not achieve your stated sample size - this should be included in your Discussion as a limitation with an acknowledgement that you may have been underpowered to detect differences in your outcomes between cases and controls.

We have added power calculations to the supplementary material and referenced them in the methods section. We have also included this problem in the discussion as a limitation.

In your methodology, there is likely to be an effect of clustering at the level of the individual GP as well as the level of the GP practice, on outcomes. Have you accounted for clustering (I could not see this)? If not, why not?

There are technical limitations to the ability to cluster in present software. We did investigate residual intraclass correlations from OLS versions of the models to see where clustering (at individual-unindexed patient, practice and network level) most affected results (and we did not have access to data on individual GPs). This identified that almost all the impact was at the individual level, rather than the practice level. Currently, Stata cannot run multilevel negative binomial models with weights included. Thus, we relied on negative binomial models with cluster robust standard errors at the individual level and included covariates to control for the health network. The use of weights is fundamental to the genetic matching and entropy balancing methods used in the study, and so multilevel negative binomial modelling was not possible. We have made reference to the investigations of intraclass correlation in the methods text and have added these investigations to supplementary material.

How many hospitals were included in your analyses? For international readers, can you clarify that you will have captured all relevant hospitals for your patient group? Will patients have gone to hospitals outside the study area, as they may well do in Australia?

Patients do indeed go to hospitals outside the study area (which is geographically small, a London Borough). The data links all NHS hospital episodes from England to the patient. Thus, all NHS hospitals in England are included in the data. Patients were highly unlikely to go to hospitals outside of England. 79% of inpatient admissions in the original raw data were in two hospitals, 90% at seven hospitals, 99% in 38 hospitals, even though there were 198 hospitals with at least one admission recorded. We have referred to all hospitals in England being included in the description of study endpoints.

3. Discussion: there are several other limitations that need to be addressed, including:

- did you take into account duration of IC that the cases received (as per your protocol)? Could it be that there was no effect seen on admissions because of insufficient IC dose and duration?

Yes, we included an offset term in the models to account for exposure/duration of IC, which could have been less than one year due to death during the follow-up period (fewer than 5% of integrated care patients died during the year). Otherwise, all patients in the integrated care group experienced one year of being enrolled on the integrated care pathway. The study looks at the effect of the programme as a whole on patient healthcare utilisation, and so there was no ‘dose’ per se, other than

the duration of enrolment. We have added discussion of the relatively short follow-up period as a limitation of the study. We have also added some emphasis to the exposure in the statistical approach of the methods section.

-please also mention your power.

We have added in reference to the original protocol power calculation and have added additional power calculations in the supplementary material. We have also added in a reference to power in the strengths and weaknesses of the discussion.

- how did the 30% of IC cases with non-matched controls differ from the 70% with matched controls on baseline variables? Was there any evidence of differences in potential confounding variables?

Good question. We have added some supplementary tables and referred to them in the text. These tables compare the means and standardised differences of those patients in the two groups who were matched and unmatched. The main issue remained related to the risk scores of the patients. The highest risk integrated care patients could not be matched; similarly, only the highest risk control patients were useful as matches.

Can you also include in the Discussion your:

- future research recommendations (eg longer follow up, larger sample, subgroup analyses of patients with/without a mental health condition, inclusion of other outcomes including re-admission rates and costs), as proposed in your protocol?

We have checked the protocol and included further discussion of future research recommendations.

- policy recommendations - what should commissioners do now, based on your findings?

We have added some policy recommendations to the conclusions based on the findings of this research.

Reviewer: 4

Reviewer Name: Ian Duncan

Institution and Country: Dept. of Statistics & Applied Probability - University of California Santa Barbara - U.S.A.

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

Please leave your comments for the authors below

Very well-researched and presented paper. I have no suggestions or concerns of substance.

I like this paper a lot for two reasons:

1. The authors are willing to publish a paper without positive outcomes. Given publication bias, this happens rarely, particularly with this type of intervention, and so makes a considerable contribution to the literature.

2. The statistical analysis is unusually robust and described in some detail. Indeed the analytical portion of the paper alone makes this paper worthwhile reading.

Many thanks.

If I have a quibble it is the length of the paper. It could well be a chapter in a book. The introduction, for example, runs 2-1/2 pages. The introduction includes an interesting history of care management efforts. I will let the editor decide whether this is appropriate for the journal. I assume that the attachment – the protocol "Evaluation of the Integrated Care Program ….Waltham Forest et al."- will not form part of the actual publication?

Other reviewers also feel that the paper is too long. We have moved some of the methods detail to the supplementary material. The protocol will be included in the supplementary material.

Some questions arose as I reviewed the paper.

1. The endpoints of the study make sense and those that I would expect (A&E, bed days etc.). I am a bit surprised by the increase in elective admissions and would appreciate more detail, such as: what admissions are considered elective vs. non-elective? In the U.S. we consider elective admissions to be those for elective procedures, i.e. non-emergency, and sometimes scheduled admissions (e.g. for repair of a hip replacement). An increase in truly elective admissions suggests some element of induced demand, which is the reverse of what one would expect with this type of management program. Perhaps the increased contact with the medical system is responsible for inducing demand, an hypothesis that the authors raise toward the end of the paper.

We have amended the description of an elective admission in the UK. We believe the most likely explanation for the increase in electives in Tower Hamlets is that there is a considerable amount of unmet need. Tower Hamlets is a very deprived part of London, and has a high level of poverty and poor health. We believe the increase was likely due to valid healthcare needs being identified and treated in the integrated care group.

2. I have not studied the QAdmissions risk score paper recently, but will try to find time to do so. In thinking about what could be driving the counter-intuitive results, I wonder whether this has to do with the risk score methodology? In the U.S. we observe fairly significant regression to the mean in terms of risk scores in elderly populations. One reason for this is that the risk score is diagnosis driven, and hospital records contain much more detailed diagnoses, so a patient who has had a recent admission is more likely to have a higher risk score (and not to have one the following year if not re-admitted). Regression does not seem to be the issue with this study, since elective admissions actually increased, but I am left wondering about the calculation and role of the risk score.

Historically, reliance on hospital data has been a concern with these kinds of risk scores (see Billings et al. 2006). In this case, we do not think regression to the mean is a problem, as the risk score was developed to primarily use data sourced from general practice records (demographics, lifestyle

variables, chronic diseases, prescribed medication, clinical values, and lab test results – see Hippisley-Cox & Coupland, 2015) alongside information on emergency admissions in the last year. Also, as you identify, we would not expect these results if regression to the mean was occurring.

3. The authors have used Kaplan-Meier analysis to check the survival of the intervention and control groups, which is appropriate. Survival, however, raises a different question: do the authors have data to explore survival longitudinally? Continuing to puzzle over the results and the higher use of elective admissions, I wonder whether this in any way contributes to longer survival of the intervention population? The authors would likely need a few years of longitudinal data to assess survival. Thinking about analysis of longitudinal data I wonder also what the pattern of risk scores looks like, post-program? I am not suggesting that the authors add to the analysis, but this could be a topic for a subsequent paper.

Yes, we agree that this would be an interesting avenue of research, and we will consider this for future studies.

Reviewer: 5

Reviewer Name: Anup Karan

Institution and Country: Indian Institute of Public Health-Delhi (IIPHD), Public Health Foundation of India (PHFI)

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

Please leave your comments for the authors below

Effect of an integrated care pathway on use of primary and secondary health care by patients at high-risk of emergency inpatient admission: a matched control cohort study in Tower Hamlets

Comments:

This study estimates effect of an integrated care pathway on primary and secondary healthcare use by high-risk of emergency inpatient in Tower Hamlets, a deprived, inner city London borough. The study uses real-time person-level administrative data and estimates effects of a range of interventions, such as case management; support with self-care; discharge support; enhanced care coordination; and specialist input in community settings on endpoints namely numbers of emergency inpatient admissions and elective inpatient admissions, inpatient bed days, accident and emergency attendances, outpatient attendances and general practitioner contacts in the year after enrolment. The study uses improved techniques of matching and create comparable sample before applying mean difference tests and difference-in-differences regression analysis. The methods also claim to control roles of unobserved covariates. On the whole, the methods in the study looks seemingly robust leading to unbiased results. I don’t have any major comments on the paper. However, I have a few minor observations which authors may like to address. May main comments are as follows:

Minor Comments:

1. After complete matching the number of observation left for final analysis are 1,720 in treatment group and 2,867 in matched control group. This reflects a large reduction in the original sample of 10,411 in the treatment and 29,885. Even after making all ineligible participants the treatment and control samples were 2,459 and 97,040 respectively before matching. Matching left very small sample. I just to flag that this might create some kind of skewed sample and may have implications on the generalibility of the results. I would like to know reply of authors on this issue else this can be mentioned in the limitation of the study.

Yes, this is a good point. We have added some supplementary tables and referred to them in the text. These tables compare the means and standardised differences of those patients in the two groups who were matched and unmatched. The main issue remained related to the risk scores of the patients. The highest risk integrated care patients could not be matched. Only the highest risk control patients were useful as matches.

2. The entire method section of the paper, although looks robust, looks little crowded. I would like to send some materials from the method section into supplementary material. For instance, the matching algorithm presented in Box 1 may be shifted to supplementary material. Authors may also like to condense the method section by putting some other details in supplementary materials.

Other reviewers have also made this point, and we agree with you. We have moved some of the detail on the methods (Box 1) to the supplementary material.

3. All the figures do not carry the Figure numbers and the required titles as to make one-to-one correspondence with the text in the manuscript.

Thank you, checked and corrected.

4. I think the main results of the study are in Table 3 and Table 4. I guess, putting Table 1 and Table 2 and the related descriptions in the supplementary file will improve the readability of the paper. The main manuscript may include only brief results of the matching.

Considering the online nature of BMJ Open, we feel this is less relevant, as tables and figures are presented in expandable web GUI ‘accordion’ sections and all sections are easily navigable.

5. The section on ‘statement of findings’ just consolidate the main findings. I would suggest authors to provide some probable explanations of almost no effects of the interventions on the endpoints. Also, can authors provide their opinion if such interventions are expected to affects the endpoints in general or not?

We have amended the discussion and conclusions to explain what these interventions aim to achieve (reductions in expensive hospital admissions) and how the evidence (including our study) suggests that they do not consistently meet these objectives, at least early-on in implementation. The main message to policy makers is that they need to be realistic as to what integrated care pathways can achieve, at least in the short term.

VERSION 2 – REVIEW

REVIEWER	Rachael Hunter UCL, UK
REVIEW RETURNED	07-Mar-2019

GENERAL COMMENTS	This is an excellent and well written paper. The authors have adequately identified the limitations of what is a challenging study.

REVIEWER	Harriet Hiscock Murdoch Children's Research Institute, Australia
REVIEW RETURNED	26-Feb-2019

GENERAL COMMENTS	I am happy with the authors' revisions to this paper.

REVIEWER	Ian Duncan Dept. of Statistics & Applied Probability University of California Santa Barbara
REVIEW RETURNED	18-Mar-2019

GENERAL COMMENTS	The format of the revised paper is much improved; it is now tighter and easier to follow. I continue to view this as an important contribution to the literature (both in terms of results and methodology) and I look forward to its publication.

REVIEWER	Anup Karan Public Health Foundation of India, India
REVIEW RETURNED	18-Mar-2019

GENERAL COMMENTS	Authors have revised the manuscript significantly and I have no major comments. However, I still don't see numbering in the figures presented at the end of the manuscript. I also don't see point-wise rebuttal by authors to my earlier minor comments, although I can see most of these comments are addressed in the revised manuscript.