PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Population-level impact of diabetes integrated care on commissioner payments for inpatient care among people with type 2 diabetes in Cambridgeshire: a post-intervention cohort follow up study
AUTHORS	Yu, Dahai; Yang, Wei; cai, yamei; Zhao, Zhanzheng; Simmons, David

VERSION 1 – REVIEW

REVIEWER	Inna Feldman Department of Public Health and Caring Sciences Uppsala University, Sweden
REVIEW RETURNED	02-Apr-2017

GENERAL COMMENTS	This is a very interesting work aims to estimate the outcomes of a re-structured health care services for diabetes patients at population level. The topic is very important since health care provider do many efforts to optimize health care services in order to improve health outcomes and decrease societal costs. That is why any development of evaluation methodology of such kind of intervention contribute to the literature. However, there are some areas of the manuscript that would benefit from revision as noted below. Overall It is necessary to define what kind of outcomes are subject for the analysis: societal costs/ health care costs/inpatient care costs/health gain. The clear aim of the study should be presented. 1) The title of the manuscript " Population-level impact of diabetes integrated care on payments for inpatient care among people with type 2 diabetes in Cambridgeshire" is confusing. If you say "payment" – what do you mean? From the manuscript, the reader can understand that the authors mean "tariff". Payment can be also "out of pocket" costs, etc. I suggest to use other terms, such as "health care costs for impatient care", "societal costs for impatient care"

	2) In the abstract: a. Line 9: "…on inpatient payments (tariff)" – also unclear, because in different countries it could be different systems for inpatient payments. "Health care costs" would be more appropriate. b. Line 16-18: "The area between the two overlapping distribution curves of inpatient cost at baseline and follow-up (at 3 years) was used to estimate the impact of integrated care on inpatient payments on a population level" -. Practically you estimated the changes in distribution curves by estimating the area between the two overlapping ….and so on. These sentences should be re-formulated c. Line 22: "…patients with diabetes". You do not use the individual data and you cannot say "patients with diabetes". Do you mean "diabetes population", "the diabetes population admitted to inpatient care"? 3) Strengths and limitations of this study a. "With application of this novel method, this study found that the integrated diabetes care was not associated with substantially reduced inpatient payments" – This is neither strength no limitation b. "The data used in this study depended on the completeness of the coding of diabetes, although there being no systematic change in coding over this time period" – this statement is unclear 4) Introduction a. The aim of the study is not presented. What questing are you going to answer to fulfill the existing gap? 5) Method a. Reference is needed line 51 (ECF: 2009 population 160,000, diabetes population 7,790) b. Line 42-49: "However when using real data to estimate parameters for two normal distributions, it is unlikely that the two curves have the same standard deviation. In our case, the two curves will have crossover points. To overcome this, we have modified the Sarkadi's method as described in supplemental technical appendix". This statement is not correct. In the paper "Sarkadi et.al, 2004" A novel approach used outcome distribution curves to estimate the population-level impact of a public health intervention", there is no assumptions that the two curves have the same standard deviation, contrariwise, the examples were calculated using different standard deviation, see fig 2 page 789. Even if you have two crossover points, the method described in Sarkadi et al is completely applicable and the authors did it correctly presenting in technical appendix. Please, clarify what do you men with the modification of the method. 6) Results and discussion The authors confuse "health gain" and "impatient payment". Health gain are not the same as "reduced inpatient payments". Why the reduction in impatient payment can be regarded as health gain? Some hypotheses behind that?

REVIEWER	Krish Nirantharakumar Institute of Applied Health Research College of Medical and Dental Sciences University of Birmingham UK I am a member of the Diabetes UK Clinical Scientific Group for Inpatient Care.
REVIEW RETURNED	17-May-2017

GENERAL COMMENTS	The authors have attempted to quantify the health gain achieved from an integrated diabetes service implemented in East Cambridgeshire and Fenland. Authors acknowledge the difficulties posed in evaluating such services and have utilised a novel method to assess the benefit of the service, though the assessment is limited by utilising inpatient cost. Introduction The trial referred in the introduction (reference 7) is incorrectly interpreted: "showed no impact on outcomes". Only the combined primary end point was insignificant but secondary outcomes such as glycaemic control and cholesterol control were significant. However the actual incremental benefit seems minimal. May be rephrase the sentence for clarity. No clear justification or reference stated why cluster RCT are not appropriate for evaluation of such services. Methods Please clarify what is meant by “There is no major hospital, falling within 4 major hospital catchment areas”. It is unclear why only impact on inpatient cost is assessed. If the aim was to assess the inpatient cost only then the conclusion the care model is not comparatively beneficial cannot be derived from the data presented. I am unable to critique the analysis as I do not have the expertise or knowledge of the techniques used. Results Findings suggest in one age group there was beneficial financial impact, however in Greater Cambridgeshire there was beneficial effect seen in both age groups under consideration. Are the authors aware of any initiative within the hospital or externally in Greater Cambridgeshire that might have resulted in such an observation? Discussion The authors say the goal of the integrated care was mostly aimed at reducing referral to outpatient setting. If that is the case why has the modelling only focused on inpatient care costs?

REVIEWER	Dr Lauren Rodgers University of Exeter Medical School, UK
REVIEW RETURNED	19-Jul-2017

GENERAL COMMENTS	This paper reviews the impact of integrated care on inpatient tariffs in rural England. There are clarifications in the analysis and reporting of the data which should be addressed prior to publication. Specific comments 1. Page 3. Bullet 2 – should be “negative values were raised”. Bullet 4 does not make sense. 2. Methods/Results/Table 1/Table 2. No N is provided for any of the regions or age groups. Sample size is of huge importance when analysing and interpreting results. This omission must be corrected. 3. Methods paragraph 2, page 5. Are code lists to be published? Reference error on line 36. 4. Methods paragraph 3, page 5. Why are baseline and follow-up sampled separately in the bootstrap? Surely they are linked data? Please justify. Does the second sentence imply that there are different numbers of baseline and follow-up data here? Please indicate if this is the case and provide details on how much data are missing and why. 5. Results page 6. What are patients grouped above and below 70? Do the results differ for different cut-offs? Reference to Figure-2. It is not clear if this refers to Figure 2 in the supplementary information or the figure on page 14. Line 50 typo “iss”. 6. Results Table 2. Results in table are not presented as percentages as per the label. Please reduce the decimal places once converted to percentages to two at most. 7. Supplementary Figure 2. The label mentions percentage of people but histograms are of frequency. What does the dashed line represent in the plots? Similar, why is there a line at 0? 8. Discussion page 7. It is not clear how the results translate into costs and investment amounts. What would constitute a ‘good’ outcome compared to the results obtained? 9. Page 14. Figure is not labelled and does not have a caption. It is unclear what the different plots in the figure represent. What is the difference between them? Is this plot referenced in the text?

REVIEWER	Kenneth J. Wilkins, PhD Biostatistics Program, Office of the Director, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, USA
REVIEW RETURNED	26-Jul-2017

GENERAL COMMENTS	The authors must be commended for not only innovating an integrated care approach for type 2 diabetes patients, but also seeking an innovative population-level approach for inference about the relative impact of the Diabetes Integrated Care Initiative (DICI) on inpatient care cost. This is an important estimate to report in a reliable manner to decision-makers considering integrated care -- especially so in recent years, given the increasing majority that inpatient costs take among total diabetes-associated expenditures in the NHS (using Kanavos, van den Aardweg & Schurer, 2012 as a reference). With this context, the authors are to be encouraged to make their findings' conclusions more reproducible via major (yet crucial) revisions to their analysis, so their design and integrated care initiative's efforts may come to fully reliable use by stakeholders. I outline these needed changes below, after outlining key references. References ---------- For more on UK's NHS cost burden getting substantial contribution (>60%) from inpatient costs (albeit for both types of diabetes), placed in context within other EU-affiliated nations, see Kanavos, van den Aardweg & Schurer, 2012 report from the London School of Economics: http://www.lse.ac.uk/LSEHealthAndSocialCare/research/LSEHealth/MTRG/LSEDiabetesReport26Jan2012.pdf For more on adapting the innovative analysis approach to the counterfactual question of whether the DICI's changes in East Cambridgeshire and the Fenlands reduced inpatient cost relative to what it would have been in the absence of DICI (as estimated by followup in 'control' portions of the catchment area), refer to the invited commentary by Maarten Bijlsma on the original population-level impact method proposed by Sarkadi and colleagues: http://dx.doi.org/10.1016/j.jclinepi.2014.06.005 The G-computation approach that he mentions -- among other select causal inference methods -- are the truly appropriate adaptations of Sarkadi et al's approach to yield estimates helpful for policymakers considering the impact of potential system-wide 'interventions' such as DICI; the original method's authors endorse this as well (http://dx.doi.org/10.1016/j.jclinepi.2014.06.006). For more on increasing statistical power for within-system changes in inpatient costs, by incorporating correlation between baseline and followup cost distributions, consult textbooks covering multivariate data methods for Gaussian and other generalized linear model outcomes (e.g., Fitzmaurice, Laird & Ware on longitudinal data, Goldstein on multilevel data, among others)

	Major revisions needed a. replace the poorly-approximating normal/Gaussian distribution for inpatient cost with one more appropriate to the right-skewed costs often cited in this literature (e.g., gamma, log-normal or some finite mixture thereof), readily implemented using R software as done in the current writeup; unlike Sarkadi et al's Eyberg Childe Behavior Inventory, your outcome has not been summed or aggregated in any manner that helps you appeal to law of large number approximations (at least as currently described). The medians and interquartile ranges reported in Table 1 seem consistent with right-skewed cost distributions, afterall. b. do not hamper your statistical power by failing to leverage the likely positive dependence between inpatient cost distributions at baseline and followup; even if your available data in NHS's Secondary Uses Service (SUS) portal does not allow individual-level, some aggregate measures of repeat-inpatient-admissions in a given time period would provide a range of plausible values to consider when reporting estimates. To underscore why this is crucial, bear in mind that the apparent lack of impact may be in fact due to underpowered statistical analysis -- one that implicitly assumes an untenable proposition for a 'closed' health care system: patients never require inpatient care again after having it once, in other words, costs at the two time-points are completely independent of one another. At the very least, you will provide a range for other systems' decision-makers to consider relative to their populations' propensity for repeating inpatient care in a similar time period. c. recognize that the stated objectives' use of the term 'impact' necessitates some form of causal inference, as outlined in the invited commentary on Sarkedi et al's population-level impact analysis approach; authors need to explicitly acknowledge that current analysis (even after meeting revisions a & b) may still yield findings subject to confounding bias -- the 'impact' may only reflect unmeasured changes in the DICI and 'control' areas respectively, rather that DICI itself as the DICI care model was not randomly assigned...you want to present estimates that have a causal interpretation (at least under the unverifiable assumption of no unmeasured confounders). Less crucial revisions Meet health economics reporting guidelines as well as observational study guidelines. While STROBE is a helpful start to transparently reporting features of study per emerging standards, the authors should also (instead?) report their responses to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement checklist. As stated in BMJ publication on CHEERS, Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement [https://www.ispor.org/TaskForces/documents/CHEERS-Statement.pdf] the type of study submitted for publication may be a "Cost consequences analysis" -- it examines costs and consequences without attempting to isolate a single consequence or aggregate consequences into a single measure, as it simply is quantifying how cost differed from what it otherwise would have been.

	The citation for the CHEERS Statement is in BMJ as well by the authors: Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, on behalf of the CHEERS Task Force.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Inna Feldman
Institution and Country: Department of Public Health and Caring Sciences, Uppsala University, Sweden
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Comments attached

Comment: This is a very interesting work aims to estimate the outcomes of a re-structured health care services for diabetes patients at population level. The topic is very important since health care provider do many efforts to optimize health care services in order to improve health outcomes and decrease societal costs. That is why any development of evaluation methodology of such kind of intervention contribute to the literature.

However, there are some areas of the manuscript that would benefit from revision as noted below.
Overall
It is necessary to define what kind of outcomes are subject for the analysis: societal costs/health care costs/inpatient care costs/health gain. The clear aim of the study should be presented.

Response: Thank you for the suggestion. We declared the objective of this study in the abstract section (page-2) as "assessed the impact of introducing a community service led diabetes integrated care programme on inpatient payments (tariff) in rural England": this is the component of healthcare costs which are paid for by the local NHS commissioners.
We have changed the objective in the abstract to:
Few studies have estimated the effect of diabetes integrated care at a population level. We have assessed the impact of introducing a community service led diabetes integrated care programme on commissioner payments (tariff) for inpatient care in rural England.
We have also changed the introduction last section to:
"Under the English National Health Service (NHS), public inpatient care is paid for from taxation through local commissioners. These payments do not generally cover the hospital costs of inpatients with diabetes [10], but can provide an NHS commissioner perspective that reflects both acuity and complexity, beyond eg length of stay. We have now used the Sarkadi approach to assess whether any changes in population based commissioner inpatient payment data occurred during a diabetes integrated care intervention by viewing the level and distribution of commissioner inpatient payments in the population as the unit of interest"

1) The title of the manuscript “Population-level impact of diabetes integrated care on payments for inpatient care among people with type 2 diabetes in Cambridgeshire” is confusing. If you say “payment” – what do you mean? From the manuscript, the reader can understand that the authors mean “tariff”. Payment can be also “out of pocket” costs, etc. I suggest to use other terms, such as “health care costs for impatient care”, “societal costs for impatient care”

Response: We thank the reviewer for alerting us to this possible confusion. These are not costs as we have previously shown that Government NHS payments for inpatients with diabetes do not cover the associated inpatient costs. We have not used the word tariff as this could also mean user or commissioner payment. We have therefore changed this to ‘commissioner’ payments and defined this term in the text. Change to:

“Population-level impact of diabetes integrated care on commissioner payments for inpatient care among people with type 2 diabetes in Cambridgeshire”

2) In the abstract:
a. Line 9: “-on inpatient payments (tariff)” – also unclear, because in different countries it could be different systems for inpatient payments. “Health care costs” would be more appropriate.

Response: As indicated, these are not costs, but payments. We have now changed the objectives to: “Few studies have estimated the effect of diabetes integrated care at a population level. We have assessed the impact of introducing a community service led diabetes integrated care programme on commissioner payments (tariff) for inpatient care in rural England.”

b. Line 16-18: “The area between the two overlapping distribution curves of inpatient cost at baseline and follow-up (at 3 years) was used to estimate the impact of integrated care on inpatient payments on a population level” -. Practically you estimated the changes in distribution curves by estimating the area between the two overlapping -.and so on. These sentences should be reformulated

Response: Apologies, but we do not understand this comment. The area under the curve here is the sum of the inpatient payments over the time period (calculus). The area under the curve is independent of the distribution curve itself-we could have had a greater or lesser peak, with a lesser or greater range, but the sum of payments would be the same. We hope this helps. We have replaced the section as follows and hope this helps:
“Commissioner data was provided by the local authority. The difference in area between the two overlapping distribution curves of inpatient payments at baseline and follow-up (at 3 years) was used to estimate the effect of integrated care on commissioner inpatient payments on a population level.”

c. Line 22: “-patients with diabetes”. You do not use the individual data and you cannot say “patients with diabetes”. Do you mean “diabetes population”, “the diabetes population admitted to inpatient care”?

Response: The revision has been made as suggested.

3) Strengths and limitations of this study
a. “With application of this novel method, this study found that the integrated diabetes care was not associated with substantially reduced inpatient payments” – This is neither strength no limitation

Response: The statement has been removed as suggested.

b. “The data used in this study depended on the completeness of the coding of diabetes, although there being no systematic change in coding over this time period” – this statement is unclear

Response: The payment data used in this study is electronic health record (HER) data, the quality of which relies mainly on the completeness of the coding. In England, the completeness of coding for diabetes has generally improved since diabetes QOF indicators were introduced in 2004. In most studies using HER data, the coding issue is usually clarified as a potential systematic error. In this study period, there was no systematic change introduced, so this statement was made to clarify the situation. We have changed the statement to “The data used in this study depended upon the completeness of the coding for diabetes in the GP records. The impact of this potential ascertainment bias should have been steady as no systematic change in coding should have occurred over this time period”

4) Introduction
a. The aim of the study is not presented. What questing are you going to answer to fulfill the existing gap?

Response: We have amended the end of the last paragraph in the introduction which hopefully clarifies the issue:
“Under the English National Health Service (NHS), public inpatient care is paid for from taxation through local commissioners. These payments do not generally cover the hospital costs of inpatients with diabetes [10], but can provide an NHS commissioner perspective that reflects both acuity and complexity, beyond eg length of stay. We have now used the Sarkadi approach to assess whether any changes in population based commissioner inpatient payment data occurred during a diabetes integrated care intervention by viewing the level and distribution of commissioner inpatient payments in the population as the unit of interest”

5) Method
a. Reference is needed line 51 (ECF: 2009 population 160,000, diabetes population 7,790)

Response: The reference has been added as suggested.

b. Line 42-49: “However when using real data to estimate parameters for two normal distributions, it is unlikely that the two curves have the same standard deviation. In our case, the two curves will have crossover points. To overcome this, we have modified the Sarkadi’s method as described in supplemental technical appendix”. This statement is not correct. In the paper “Sarkadi et.al, 2004” A novel approach used outcome distribution curves to estimate the population level impact of a public health intervention”, there is no assumptions that the two curves have the same standard deviation, contrariwise, the examples were calculated using different standard deviation, see fig 2 page 789. Even if you have two crossover points, the method described in Sarkadi et al is completely applicable and the authors did it correctly presenting in technical appendix. Please, clarify what do you men with the modification of the method.

Response: Thanks for pointing this out. I agree with the reviewer that in Sarkadi’s paper, they have shown examples for the case when the two curves have different standard deviations (SD). However, the method only works when the follow-up group have lower mean, and either higher SD or the same SD. Our modification was to address the case when the follow-up group have lower mean, but lower SD also. In this case, there will be two crossover points, both of which are higher than the follow-up mean. Our modification corresponds to choosing the lambda value as the smaller of the two solutions of (A1:3) from the technical appendix, when the SD is lower in the follow up group. This is different from the other case when SD higher is in the follow-up group, the larger of the two solutions for lambda should be chosen. This was not explicitly described in in Sarkadi et al.

We have changed the descriptions for the modification to make this clear.

6) Results and discussion
The authors confuse “health gain” and “impatient payment”. Health gain are not the same as "reduced inpatient payments". Why the reduction in impatient payment can be regarded as health gain? Some hypotheses behind that?

Response: Reduction in healthcare payments under the NHS allows the liberated public funds to be used elsewhere to achieve a gain in health. We now specify this in the methods:
The reduction in commissioner payments is seen as a ‘health gain’, as under the NHS, such liberation of public funds can be used elsewhere to achieve a gain in health.

Reviewer: 2
Reviewer Name: Krish Nirantharakumar
Institution and Country: Institute of Applied Health Research, College of Medical and Dental Sciences, University of Birmingham, UK
Please state any competing interests or state ‘None declared’: I am a member of the Diabetes UK Clinical Scientific Group for Inpatient Care.

Please leave your comments for the authors below
The authors have attempted to quantify the health gain achieved from an integrated diabetes service implemented in East Cambridgeshire and Fenland. Authors acknowledge the difficulties posed in evaluating such services and have utilised a novel method to assess the benefit of the service, though the assessment is limited by utilising inpatient cost.
Introduction
The trial referred in the introduction (reference 7) is incorrectly interpreted: "showed no impact on outcomes". Only the combined primary end point was insignificant but secondary outcomes such as glycaemic control and cholesterol control were significant. However the actual incremental benefit seems minimal. May be rephrase the sentence for clarity.
Response: The statement has been changed as “One randomised trial of an intermediate care service achieved minimal actual incremental benefit [7]”.

No clear justification or reference stated why cluster RCT are not appropriate for evaluation of such services.
Response: Thank you, we have changed this sentence to “By their nature, randomised controlled trials are difficult to utilise when assessing the impact of a complete system change at a population level”

Methods
Please clarify what is meant by “There is no major hospital, falling within 4 major hospital catchment areas”.
Response:
The geographical area has no major hospital (with eg an emergency department, acute medical wards), as these lie in neighbouring areas. Revised to:
“There is no local major hospital (with eg an emergency department), falling within the catchment areas of 4 hospitals outside of the area”

It is unclear why only impact on inpatient cost is assessed. If the aim was to assess the inpatient cost only then the conclusion the care model is not comparatively beneficial cannot be derived from the data presented.

Response: We concurrently assessed the impact on metabolic control and hospitalisation and found no impact (reference 6). We have now obtained the tariff (commissioner payment data), which can sometimes show benefit as it reflects different components beyond hospitalisation (yes/no) and length of stay such as complexity and acuity of admissions. We have now added in the methods:
We have previously reported no impact on metabolic control or hospitalisation rates in spite of full implementation of the service (reference 6).

I am unable to critique the analysis as I do not have the expertise or knowledge of the techniques used.
No changes made

Results
Findings suggest in one age group there was beneficial financial impact, however in Greater Cambridgeshire there was beneficial effect seen in both age groups under consideration. Are the authors aware of any initiative within the hospital or externally in Greater Cambridgeshire that might have resulted in such an observation?

Response: Yes, local hospital diabetes services had continuous quality improvement programmes, however, we can not necessarily attribute such changes to this. We have added: '…although each hospital based service would have continued with its own internal service developments.'

Discussion
The authors say the goal of the integrated care was mostly aimed at reducing referral to outpatient setting. If that is the case why has the modelling only focused on inpatient care costs?

Response: Yes, we agreed that the goal could be both at inpatient and outpatient care. Due to the access of data, we could not evaluate the outpatient care in this study, we have admitted this in the Limitation section.
The goal was not reducing referrals, this was the prior philosophy, before the wider programme was introduced. We have therefore changed the text to:
This philosophy, rather than progressing to truly integrated services carried through the intervention period, albeit as part of a wider programme that included 'vertical integration' developments. It was perhaps to be expected that attempts at creating such greater 'vertical' integration in information management, clinical governance, budget and overall management were agreed but not implemented, actions more achievable within a single organization.
In addition, as above, we have already published on the wider changes (reference 6) where we described the impact on metabolic control and hospitalisation. As above, tariff data provides information on acuity and complexity beyond hospitalisation and length of stay.

Reviewer: 3
Reviewer Name: Dr Lauren Rodgers
Institution and Country: University of Exeter Medical School, UK
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
This paper reviews the impact of integrated care on inpatient tariffs in rural England. There are clarifications in the analysis and reporting of the data which should be addressed prior to publication.
Specific comments
1. Page 3. Bullet 2 – should be "negative values were raised". Bullet 4 does not make sense.
Response: We have changed the text to:
- The 'health gain' in the revised method was clearly defined with a formulated algorithm of evaluation, which broadened the utilization scenarios especially when negative values were raised.

2. Methods/Results/Table 1/Table 2. No N is provided for any of the regions or age groups. Sample size is of huge importance when analysing and interpreting results. This omission must be corrected.

Response: Due to the limitation in data access, there was no personal identification to identify the repeated records occurred by the same person. As all analyses were record based study. The number of hospitalised records were listed in the supplemental table 1.
We now include the numbers within table 1 and dropped supplemental table 1

3. Methods paragraph 2, page 5. Are code lists to be published? Reference error on line 36.

Response: The code is available for review and will be shared to readers upon request after publication. The reference error has been corrected.

4. Methods paragraph 3, page 5. Why are baseline and follow-up sampled separately in the bootstrap? Surely they are linked data? Please justify. Does the second sentence imply that there are different numbers of baseline and follow-up data here? Please indicate if this is the case and provide details on how much data are missing and why.

Response: We thank the referee for highlighting this area of possible confusion. As previously described, the records are based upon anonymised, rather than individual, data. As a result, it was not possible to match samples for the baseline and follow-up analyses. However, using the method in the manuscript, the population-wide changes could be evaluated based upon the hypothesis that the baseline and follow-up inpatient payments occurred within the same diabetes population.
We have clarified this in the Figure legend for Supplemental Figure 2 as:
"The health payment gain (impact) was defined as percentage of people with type 2 diabetes and hospital admission having reduced heath payments after 3 years of diabetes integrated care at a population level. Distributions of the impact were approximated by bootstrapping. Subjects and their associated payments were selected by random resampling with replacement from the original data 10000 times, and the impact was calculated in each resampled dataset. The dashed red line shows the impact in the original data. The bootstrap p value was calculated by comparing the impact estimated in the resampled data to 0 (indicated by the solid red line; null hypothesis H0: impact<=0, and alternative hypothesis H1: impact>0; P = [Percentage with impact <= 0])".

5. Results page 6. What are patients grouped above and below 70? Do the results differ for different cut-offs? Reference to Figure-2. It is not clear if this refers to Figure 2 in the supplementary information or the figure on page 14. Line 50 typo "iss".

Response: Age as a common, and significant confounder, was tackled by stratification analysis in this study. 70 was the median of patients generated the inpatient care record, we now highlight this as the reason for this age for dichotomisation. Sorry for the typo. There is no Figure-2 in the main text. The age-stratified results were both refer to Figure-1 and supplemental Figure-2. The typo 'iss' has been corrected to 'is' as suggested. We have added a sentence in the methods:
"Age data were provided allowing analyses to be undertaken above and below the median age (70 years) to assess any related variation".

6. Results Table 2. Results in table are not presented as percentages as per the label. Please reduce the decimal places once converted to percentages to two at most.

Response: The revision has been made as suggested.

7. Supplementary Figure 2. The label mentions percentage of people but histograms are of frequency. What does the dashed line represent in the plots? Similar, why is there a line at 0?

Response: The figure legend of supplemental figure 2 has been updated as below “The health payment gain (impact) was defined as percentage of people with type 2 diabetes and hospital admission having reduced heath payments after 3 years of diabetes integrated care at a population level. Distributions of the impact were approximated by bootstrapping. Subjects and their associated payments were selected by random resampling with replacement from the original data 10000 times, and the impact was calculated in each resampled dataset. The dashed red line shows the impact in the original data. The bootstrap p value was calculated by comparing the impact estimated in the resampled data to 0 (indicated by the solid red line; null hypothesis H0: impact<=0, and alternative hypothesis H1: impact>0; P = [Percentage with impact <= 0])”.

8. Discussion page 7. It is not clear how the results translate into costs and investment amounts. What would constitute a ‘good’ outcome compared to the results obtained?

Response: We have now described how commissioner savings under the NHS can be used elsewhere to benefit the local health economy. We have stated this on page 10 as “The latest changes in commissioning in the English NHS, with emphasis on the need to consider ‘Any qualified Provider’ in service delivery, and associated market procurement approaches, could well impair the quality of diabetes care while increasing overall cost, if the experience here is reproduced elsewhere. ”.

9. Page 14. Figure is not labelled and does not have a caption. It is unclear what the different plots in the figure represent. What is the difference between them? Is this plot referenced in the text?

Response: Figure-1 was uploaded as a separate file and appeared as an unlabelled figure on page 14. The figure legend of figure was integrated following the main text.

Reviewer: 4
Reviewer Name: Kenneth J. Wilkins, PhD
Institution and Country: Biostatistics Program, Office of the Director, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, USA
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Comment: The authors must be commended for not only innovating an integrated care approach for type 2 diabetes patients, but also seeking an innovative population-level approach for inference about the relative impact of the Diabetes Integrated Care Initiative (DICI) on inpatient care cost.

Response: This is an important estimate to report in a reliable manner to decision-makers considering integrated care -- especially so in recent years, given the increasing majority that inpatient costs take among total diabetes-associated expenditures in the NHS (using Kanavos, van den Aardweg & Schurer, 2012 as a reference).

With this context, the authors are to be encouraged to make their findings' conclusions more reproducible via major (yet crucial) revisions to their analysis, so their design and integrated care initiative's efforts may come to fully reliable use by stakeholders. I outline these needed changes below, after outlining key references.
Response: We thank the referee for his comments.
References

Comment: For more on UK's NHS cost burden getting substantial contribution (>60%) from inpatient costs (albeit for both types of diabetes), placed in context within other EU-affiliated nations, see Kanavos, van den Aardweg & Schurer, 2012 report from the London School of Economics: http://www.lse.ac.uk/LSEHealthAndSocialCare/research/LSEHealth/MTRG/LSEDiabetesReport26Jan2012.pdf

Response: The reference has been added as suggested.

Comment: For more on adapting the innovative analysis approach to the counterfactual question of whether the DICI's changes in East Cambridgeshire and the Fenlands reduced inpatient cost relative to what it would have been in the absence of DICI (as estimated by followup in 'control' portions of the catchment area), refer to the invited commentary by Maarten Bijlsma on the original population-level impact method proposed by Sarkadi and colleagues:
http://dx.doi.org/10.1016/j.jclinepi.2014.06.005

Response: The reference has been added as suggested.

Comment: The G-computation approach that he mentions -- among other select causal inference methods -- are the truly appropriate adaptations of Sarkadi et al's approach to yield estimates helpful for policymakers considering the impact of potential system-wide 'interventions' such as DICI; the original method's authors endorse this as well (http://dx.doi.org/10.1016/j.jclinepi.2014.06.006).

Response: We thank the reviewer for pointing out G-computation and other related methods to control for confounding and for evaluating the impact of system-wide interventions. We have attempted to implement the G-computation approach, and this has not changed our findings. We now state this in the last paragraph of the results section. As a matter of fact, the present study has been performed under the assumption that baseline and follow-up inpatient payments occurred in the same diabetes population, and so the problem of confounding factors seems less critical in this study.

Comment: For more on increasing statistical power for within-system changes in inpatient costs, by incorporating correlation between baseline and followup cost distributions, consult textbooks covering multivariate data methods for Gaussian and other generalized linear model outcomes (e.g., Fitzmaurice, Laird & Ware on longitudinal data, Goldstein on multilevel data, among others)

Response: Unfortunately, we are restricted by data access and there are no further co-variables besides sample age available for multivariable analyses. We have admitted this in the limitations section.

Major revisions needed

a. replace the poorly-approximating normal/Gaussian distribution for inpatient cost with one more appropriate to the right-skewed costs often cited in this literature (e.g., gamma, log-normal or some finite mixture thereof), readily implemented using R software as done in the current writeup; unlike Sarkadi et al's Eyberg Childe Behavior Inventory, your outcome has not been summed or aggregated in any manner that helps you appeal to law of large number approximations (at least as currently described). The medians and interquartile ranges reported in Table 1 seem consistent with right-skewed cost distributions, afterall.

Response: Thank you for the suggestion. We have applied the gamma distribution and log-normal distribution in our data. However, there was no significant improvement in fitting the data distributions observed by applying these distributions. We have described these attempts in the last paragraph of the results section.

b. do not hamper your statistical power by failing to leverage the likely positive dependence between inpatient cost distributions at baseline and followup; even if your available data in NHS's Secondary Uses Service (SUS) portal does not allow individual-level, some aggregate measures of repeat-inpatient-admissions in a given time period would provide a range of plausible values to consider when reporting estimates. To underscore why this is crucial, bear in mind that the apparent lack of impact may be in fact due to underpowered statistical analysis -- one that implicitly assumes an untenable proposition for a 'closed' health care system: patients never require inpatient care again after having it once, in other words, costs at the two time-points are completely independent of one another. At the very least, you will provide a range for other systems' decision-makers to consider relative to their populations' propensity for repeating inpatient care in a similar time period.

Response: Thank you for this good suggestion. We have added this statement in the discussion section.

c. recognize that the stated objectives' use of the term 'impact' necessitates some form of causal inference, as outlined in the invited commentary on Sarkedi et al's population-level impact analysis approach; authors need to explicitly acknowledge that current analysis (even after meeting revisions a & b) may still yield findings subject to confounding bias -- the 'impact' may only reflect unmeasured changes in the DICI and 'control' areas respectively, rather that DICI itself as the DICI care model was not randomly assigned...you want to present estimates that have a causal interpretation (at least under the unverifiable assumption of no unmeasured confounders).

Response: Thank you for this good suggestion. We have acknowledged the existence of unmeasured confounders in the current analysis and stated in the limitation section.

Less crucial revisions

Meet health economics reporting guidelines as well as observational study guidelines.
While STROBE is a helpful start to transparently reporting features of study per emerging standards, the authors should also (instead?) report their responses to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement checklist.

As stated in BMJ publication on CHEERS, Consolidated Health Economic Evaluation Reporting Standards (CHEERS) Statement
[https://www.ispor.org/TaskForces/documents/CHEERS-Statement.pdf]
the type of study submitted for publication may be a "Cost consequences analysis" -- it examines costs and consequences without attempting to isolate a single consequence or aggregate consequences into a single measure, as it simply is quantifying how cost differed from what it otherwise would have been.
The citation for the CHEERS Statement is in BMJ as well by the authors:
Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, Augustovski F, Briggs AH, Mauskopf J, Loder E, on behalf of the CHEERS Task Force.

Response: Thank you for the suggestion. We have submitted the CHEERS checklist as suggested.

VERSION 2 – REVIEW

REVIEWER	Inna Feldman Department of Public Health and Caring Sciences Uppsala University, Sweden
REVIEW RETURNED	11-Sep-2017

GENERAL COMMENTS	The authors have clarified my entire questions. I have only one concern left. While defined the 'health gain', the authors stated on page 5-6: In our study, the 'health gain' represents the proportion of patients with reduced inpatient payments between the baseline and intervention period. The reduction in commissioner payments is seen as a 'health gain', as under the NHS, such liberation of public funds can be used elsewhere to achieve a gain health. I think that this explanation is not correct. I would suggest to explain the guess that the reduced inpatient payments reflects reduced needs in care and thus improvement in health.

REVIEWER	Krish Nirantharakumar IAHR, University of Birmingham, UK Member of the inpatient clinical scientific group of Diabetes UK Expert topic member for the NICE guidelines update on diabetes prevention and management
REVIEW RETURNED	28-Sep-2017

GENERAL COMMENTS	The limitations that were raised have been addressed or discussed.

REVIEWER	Kenneth J. Wilkins, PhD Biostatistics Program, Office of the Director, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, USA
REVIEW RETURNED	23-Sep-2017

GENERAL COMMENTS	It's encouraging to see the authors address a slew of much-needed revisions requested by the four original reviewers. All changes to this point, ranging from efforts to sharpen terminology, meet emerging standards (e.g., Consolidated Health Economic Evaluation Reporting Standards, CHEERS) or elaborate on data/study limitationshave been made in the spirit of improving the research presentation via peer review. I'm eager to see these appropriately-presented findings reach publication. <p> For this to take place, however, I do see some minor revisions required still at this stage: a. Clearly state that lack of evidence of intervention effects does not equate to evidence that intervention lacks any effect whatsoever, it merely speaks to lack of statistical power or shortfall of information that a more suitably designed study may well detect. In short, explicitly revise the article to address the following: <p> (i) Appropriately caveat your 'negative' findings as being due to data limitations...if you had longitudinal data, you would have greater power to detect between-group differences from baseline to followup (attributable to the intervention, DICI) as the fraction of individual patients who incurred "costs"/"payments" in both periods could serve as their own internal 'controls' regardless of intervention status; more bluntly, any analysis must recognize that these 'repeat-incurrers' of payments provide redundant information, such that the method of Sarkadi et al. -- as employed by Feldman et al., (when applying a bootstrap that explicitly ignores this redundancy) by resampling strictly within each timepoint -- is a mis-application of tools despite the best of intention. It cannot be ruled out that this study's findings are inconclusive simply by dint of failing to exploit additional statistical power by explicitly accounting for this (likely non-negligible) fraction who could serve as their own controls. Stated in terms that someone who presumes the null hypothesis of no difference in outcomes might, such miss-application may inadvertently increase the false positive rate: if, in fact, the naive application of Sarkadi et al's method to substantially dependent data with considerable zero-inflation does lead to an inflated Type I 'false positive' error, the observed p-values for certain NHS regions might be considered in a distinctly different light. <p> (ii) Quantify within a main-article table the degree to which your inconclusive findings change when presuming a plausible range of baseline-to-followup dependence or correlation -- while it's understandable that (without longitudinal links) you cannot empirically estimate this dependence, you can call upon similar "payment" data that you can access, estimate its measures of dependence (e.g., Pearson's correlation coefficient, with a Fisher's z-transform-based suitably-chose confidence interval) and report the extent to which your findings change under simulated bootstrapped samples that, in all other respects, match your observed data.

	(iii) Quantify the extent to which alternate distribution assumptions (e.g., gamma or log-normal) did not make improvement in fitting the data by actually reporting the estimated goodness-of-fit measures for the normal distribution versus these alternatives. It may be done in a footnote of the table or parenthetically in the narrative as space limitations dictate. <p> b. In addition to stating how findings in alternate analyses were similar to those reported in primary analyses (such as in a., gauging how inconclusive findings may change once plausibly assuming that some fraction of patients incur cost at both baseline and followup), you should quantify how different from your current (implicit) assumptions a reasonable alternative situation would have to be (as might be done in G-computation) in order for the findings to become conclusive one way (intervention worse than standard care) or the other (intervention better than standard care); this is often called a "tipping-point" analysis, as frequently employed in the missing data literature. As the data have been anonymized, apart from age, and not linked longitudinally, you need only specify the extremes of a plausible range in addition to those separate re-analyses you've already done and reported in your response as having similar results; the minor revision that's crucial (for sake of both rigor and reproducibility) is supplying these quantities in the results table or a footnote thereto. <p> c. clear up the confusion in terminology regarding "tariff"/"cost"/"payment" with an up-front definition of terms AND consistent use thereafter so that a motivated reader might avoid the confusion that the original reviewers (and to some extent, those on this second review) might encounter. <p> With these minor changes, the investigative team will have refined their manuscript to the standards of rigor and reproducibility expected of BMJ Open articles, in my opinion.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Inna Feldman
Institution and Country: Department of Public Health and Caring Sciences, Uppsala University, Sweden
Please state any competing interests or state 'None declared': None decalred

Please leave your comments for the authors below
The authors have clarified my entire questions. I have only one concern left.
While defined the 'health gain', the authors stated on page 5-6:

Comment: In our study, the 'health gain' represents the proportion of patients with reduced inpatient payments between the baseline and intervention period. The reduction in commissioner payments is seen as a 'health gain', as under the NHS, such liberation of public funds can be used elsewhere to achieve a gain health.

I think that this explanation is not correct. I would suggest to explain the guess that the reduced inpatient payments reflects reduced needs in care and thus improvement in health.

Response: Thank you for the suggestion. The statements has been revised to "The reduction in commissioner payments reflects reduced needs in care and thus improvement in health" as suggested.

Reviewer: 4
Reviewer Name: Kenneth J. Wilkins, PhD
Institution and Country: Biostatistics Program, Office of the Director, National Institute of Diabetes & Digestive & Kidney Diseases, National Institutes of Health, USA
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Comment: It's encouraging to see the authors address a slew of much-needed revisions requested by the four original reviewers. All changes to this point, ranging from efforts to sharpen terminology, meet emerging standards (e.g., Consolidated Health Economic Evaluation Reporting Standards, CHEERS) or elaborate on data/study limitation shave been made in the spirit of improving the research presentation via peer review. I'm eager to see these appropriately-presented findings reach publication.

For this to take place, however, I do see some minor revisions required still at this stage:

a. Clearly state that lack of evidence of intervention effects does not equate to evidence that intervention lacks any effect whatsoever, it merely speaks to lack of statistical power or shortfall of information that a more suitably designed study may well detect. In short, explicitly revise the article to address the following:

(i) Appropriately caveat your 'negative' findings as being due to data limitations...if you had longitudinal data, you would have greater power to detect between-group differences from baseline to followup (attributable to the intervention, DICI) as the fraction of individual patients who incurred "costs"/"payments" in both periods could serve as their own internal 'controls' regardless of intervention status; more bluntly, any analysis must recognize that these 'repeat-incurrers' of payments provide redundant information, such that the method of Sarkadi et al. -- as employed by Feldman et al., (when applying a bootstrap that explicitly ignores this redundancy) by resampling strictly within each timepoint -- is a mis-application of tools despite the best of intention. It cannot be ruled out that this study's findings are inconclusive simply by dint of failing to exploit additional statistical power by explicitly accounting for this (likely non-negligible) fraction who could serve as their own controls. Stated in terms that someone who presumes the null hypothesis of no difference in outcomes might, such miss-application may inadvertently increase the false positive rate: if, in fact, the naive application of Sarkadi et al's method to substantially dependent data with considerable zero-inflation does lead to an inflated Type I 'false positive' error, the observed p-values for certain NHS regions might be considered in a distinctly different light.

Response: Thank you for the illustration and suggestion. We have admitted this as the limitation and stated in line 10-17 on page 12 as “As a result of data access restrictions, it is not possible in this study to identify those with multiple admissions (and payments) that would provide ‘redundant information’. The application of bootstrapping ignoring such redundant information might lead to a mis-application of Sarkadi’s tool and might inadvertently increase the false positive rate: something to be taken into consideration when interpreting the findings in this study”.

(ii) Quantify within a main-article table the degree to which your inconclusive findings change when presuming a plausible range of baseline-to-followup dependence or correlation -- while it's understandable that (without longitudinal links) you cannot empirically estimate this dependence, you can call upon similar "payment" data that you can access, estimate its measures of dependence (e.g., Pearson's correlation coefficient, with a Fisher's z-transform-based suitably-chose confidence interval) and report the extent to which your findings change under simulated bootstrapped samples that, in all other respects, match your observed data.

Response: Thank you for the suggestion. Due to the limitations in the data used in this current study, we could not address this issue.

Out of interest, to address the potential problem of baseline-to-follow-up correlation, we did utilise data from another study with longitudinal links and examined the correlation of inpatient payment data correlation between baseline and follow-up. The Pearson coefficient was 0.035 (95% confidence interval: -0.196 to 0.263) between baseline and follow-up suggesting a low correlation between baseline and follow-up in that cohort. As participants in this study were derived from the sample population (same residence and same health care system), we think this would suggest a low correlated baseline-to-follow-up in our study. However, we provide this information for the reviewer and not for wider publication.

We were not allowed to use identifying variables in our study, and were therefore unable to test this correlation issue. We have acknowledged this in the limitation section.

(iii) Quantify the extent to which alternate distribution assumptions (e.g., gamma or log-normal) did not make improvement in fitting the data by actually reporting the estimated goodness-of-fit measures for the normal distribution versus these alternatives. It may be done in a footnote of the table or parenthetically in the narrative as space limitations dictate.

Response: Thanks you for the suggestion. To make the method crystal clear and help readers to fully understand the process of this method, we listed three goodness-of-fit statistics: AIC, BIC and log-likelihood over four distributions (Normal distribution, Gamma-distribution, Log-Normal distribution, and Normal distribution with log-transferred inpatient payment that was newly tested in this round of revision). The best goodness-of-fit statistics (minimum AIC, BIC and maximum log-likelihood) was chosen as the distribution to fit the data and estimate the `health gain'. After comparisons, the Normal distribution of log-transferred inpatient payment was identified to be the best-fitted distribution (see online supplemental table 1) and chosen to estimate the `health gain' (revised Table-3). Some changes were observed, but the conclusion was not reversed as the intervention still did not present significant impact on the reduced inpatient payment in the intervention areas with comparison to control areas.

We have added the statement on the comparison in line 15-20 on page 6 as "In addition to the Normal distribution originally used in Sakadi's method, other three distributions, Gamma distribution, Log-Normal distribution and Normal distribution of log-transferred payment data were attempted to fit the data. The goodness-of-fit statistics, AIC, BIC, and log-likelihood were tested over four distributions and the distribution with the minimum AIC, BIC and maximum log-likelihood was chosen as the final distribution to examine the impact" in the Methods section. The results of estimated `health gain' was also updated to the estimations based on the best-fitted distribution in line 6-16 on page 8 as "The significant `health gain' was observed both in the intervention area and control areas, especially among patients aged less than 70 years. In the intervention area, East Cambridge and Fenland, 7.69% (95 Confidence Interval (CI) 5.89-9.74%) and 2.05% (0.72 to 4.13%) of patients aged less than 70 years and aged more than 70 years, respectively had a reduced inpatient payment, compared with the population in the baseline period. In Huntingdonshire, the `health gain' was 6.90% (5.63 to 8.68%) and 4.62% (2.22 to 7.23%) among patients aged less than 70 years and patients aged more than 70 years, respectively. In Greater Cambridge, the `health gain' was 7.59% (5.63 to 9.94%) and 2.49% (1.46 to 4.58%) among patients aged less than 70 years and patients aged more than 70 years, respectively". The estimations from the original Normal distributions were re-mapped to online supplemental Table 2 to facilitate reader to understand the method.

b. In addition to stating how findings in alternate analyses were similar to those reported in primary analyses (such as in a., gauging how inconclusive findings may change once plausibly assuming that some fraction of patients incur cost at both baseline and followup), you should quantify how different from your current (implicit) assumptions a reasonable alternative situation would have to be (as might be done in G-computation) in order for the findings to become conclusive one way (intervention worse than standard care) or the other (intervention better than standard care); this is often called a "tipping-point" analysis, as frequently employed in the missing data literature. As the data have been anonymized, apart from age, and not linked longitudinally, you need only specify the extremes of a plausible range in addition to those separate re-analyses you've already done and reported in your response as having similar results; the minor revision that's crucial (for sake of both rigor and reproducibility) is supplying these quantities in the results table or a footnote thereto.

Response: Based on our current data, we could make further analysis like G-computation. We were not allowed to present other studies data and analysis results in this paper.
We did tried to apply the G-computation in another data that has information on sample correspondence and sample age. Using this data, we evaluated the potential cofounding effect for age using G computation as implemented in the "tmle" R package. To directly compare with Sarkadi's method, the residual cost was calculated, and used to estimate the intervention impact after modeling and removing the confounding effect. Before considering the effect of age, the estimated cost impact (reduced inpatient payment in financial year of 2009-2011 comparing with financial year of 2008-2009) is 0.0259 (-0.0967 to 0.1394). After explicitly modeling the confounding effect for age, the impact changed to 0.0227 (-0.0986 to 0.1363).

The bootstrap p value for the difference of estimated impacts before and after adjusting for the covariate is 0.958. This shows age is probably not a confounding effect in our study. Evaluations for other confounding factors is guaranteed with the availability of data in future studies.
But we could not added the above results into the current paper. We would acknowledge the potential confounding effect and we think the Evaluations for other confounding factors is guaranteed with the availability of data in future studies as we stated in the line 17-18 on page 12.

c. clear up the confusion in terminology regarding "tariff"/"cost"/"payment" with an up-front definition of terms AND consistent use thereafter so that a motivated reader might avoid the confusion that the original reviewers (and to some extent, those on this second review) might encounter.

Response: Thank you for the suggestion. The terminology has been unified throughout the manuscript as 'inpatient payment'.

With these minor changes, the investigative team will have refined their manuscript to the standards of rigor and reproducibility expected of BMJ Open articles, in my opinion.

Reviewer: 2
Reviewer Name: Krish Nirantharakumar
Institution and Country: IAHR, University of Birmingham, UK
Please state any competing interests or state 'None declared': Member of the inpatient clinical scientific group of Diabetes UK

Expert topic member for the NICE guidelines update on diabetes prevention and management

Please leave your comments for the authors below
The limitations that were raised have been addressed or discussed.

VERSION 3 – REVIEW

REVIEWER	Inna Feldman Uppsala University, Sweden
REVIEW RETURNED	04-Nov-2017

GENERAL COMMENTS	No comments