PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Analysis of English General Practice level data linking the levels of specific medication, service and demography to levels of glycaemic control being achieved in Type 2 Diabetes in order to improve clinical practice and outcomes.
AUTHORS	Heald, Adrian; Davies, Mark; Stedman, Mike; Livingston, Mark; Lunt, Mark; Fryer, Anthony; Gadsby, Roger

VERSION 1 – REVIEW

REVIEWER	Thomas H. Wieringa Currently without research affiliation
REVIEW RETURNED	16-Jan-2019

GENERAL COMMENTS	The topic of your research is interesting and important for improvement of the GP practices. You might consider changing "patient outcomes" as I expected the research investigating patient reported outcomes because of the use of this term. In general, the use of the English language is more than sufficient, but there are some little flaws. Please check this. If you want your method section to be reproducable, then please elaborate a little more on the statistical analyses (e.g., software used, which longitudinal analyses used, etc.).

REVIEWER	Prof. Karel Kostev IQVIA, Epidemiology, Germany
REVIEW RETURNED	28-Jan-2019

GENERAL COMMENTS	Authors present an interesting study. They have developed a statistical model to quantify opportunities for performance improvement for diabetes control within UK GP practices. They used several statistical methods and finally estimated and displayed effects of antihyperglycemic drugs and other factors on the HbA!c values. I have the following points. Major points: 1) Authors have used a multivariate linear regression. For linear regression, the depending variable like HbA1c should have continuous values like , for example, 6.5, 6.6, 6.7 ……. 10.0 and so on. On the other place they wrote that they defined optimal and poor glycemic control: “Impact on outcomes was defined by two measures:

	% of patients within GPPs at target glycaemic control (TGC) (HbA1c ≤58mmol/mol), and % of patients within GPPs with high glycaemic risk (HGR) (HbA1c >86mmol/mol" It means that the study patients can reach target or not (yes/no, 1,0). When it looks like that then the use of linear regression would be inappropriate but rather logistic regression should be used. Instead of ß-values which cannot be understandable interpreted, Odds Ratios could be shown which show very clearly the increasing or decreasing of the probability to achieve the target. 2) Each antihyperglycemic drug class and drug itself was included as separate group, for example: metformin, DPP-4i, GLP-1 and so on. However, in the diabetes therapy, especially in the case of therapy escalation, most of these drug classes are given as combinations. For example, metformin + DPP-4i, metformin + sulfonylurea, metformin + insulin and so on. The effect of metformin + insulin can be different than the effect of metformin + DPP-4i. GLP-1 + insulin can have another effect as GLP-1+metformin. When the analysis of such combinations is difficult, then authors should mention it in the limitations. Minor: - Figures have not an optimal quality and probably other colors should be selected for better presentation. On figures 3 and 4 there is not possible to read the text due to the small font size.

REVIEWER	Peter Eibich Max Planck Institute for Demographic Research, Germany
REVIEW RETURNED	30-Jan-2019

GENERAL COMMENTS	This study aims to identify factors that affect outcomes for individuals with type 2 diabetes in England. Specifically, the authors focus on the percentage of individuals with type 2 diabetes who achieve target glycaemic control (TGC) and the percentage of patients at higher glycaemic risk (HGR) at the GP practice level. They combine data from several different sources, including the National Diabetes Audit. They estimate cross-sectional and longitudinal (first-differenced) multivariate regression models to examine associations between outcomes (TGC and HGR) and factors such as prescribing, health services and demographics at the practice level. Based on these estimated regression coefficients, the authors then project the change in individuals achieving target glycaemic control or being at higher glycaemic risk if all GP practices would follow best-practice examples. The results indicate that practices using older therapy lines (e.g., sulphonylurea or insulin) show poorer outcomes, whereas practices using newer agents and a wider range of services (e.g., feet checks) perform better. Following the best-practice could lead to a considerable increase in patients with good outcomes. This is certainly a very interesting study with a good contribution – identifying practice-styles that are associated with better patient outcomes for individuals with type 2 diabetes. The paper is well-written and concise. However, I have a few concerns, in particular regarding the analysis. 1.) My main concern is that the estimates might be biased by reverse causality. You mention several times throughout the manuscript that there might be confounding, however, you interpret your findings in a causal framework, implying that patient outcomes

	would improve if prescription would change. The potential impact of confounders is not discussed (beyond acknowledging that confounding might exist), and you do not conduct any specific sensitivity analyses to address potential confounding. In particular, it seems that you only control for the percentage of patients receiving a certain medication, and not for previous treatments or the treatment line. This could bias your results, since poor outcomes would typically lead to an escalation of treatment, and thus poor outcomes might be systematically associated with different prescription patterns. For example, insulin is typically only prescribed if target glycaemic control cannot be achieved with other therapy lines, i.e., it might not actually be possible to prescribe metformin to patients that previously received insulin (as an example). Your finding that, in the cross-sectional analysis, sulphonylurea and insulin are associated with worse outcomes might be an artifact of this reverse causality, since these are typically not prescribed as first-line therapy. I would suggest to try and distinguish between therapy-lines/therapy patterns if possible, and you will need to discuss the potential bias in your results carefully. 2.) To an extent, this reverse causality bias might be mitigated by the longitudinal analysis, since in this analysis you only relate changes in prescription patterns to changes in your outcomes. However, while you briefly discuss some advantages and disadvantages of the cross-sectional and longitudinal models (mostly regarding the source of the variation), you do not seem to check which of the models is more appropriate (e.g., by using a Hausman test). This makes it difficult for the reader to assess which of the models findings are more reliable. I would suggest to provide some more extensive guidance on the differences in the interpretation of your cross-sectional and longitudinal analyses, and what this might imply for the results. 3.) In the introduction, you state that “Targeting the more effective agents is therefore fundamental to longer term affordability”. This is not necessarily clear, since the costs of these agents also differ substantially. You acknowledge in the discussion section that using newer drugs would be costly, however, the sentence also seem to suggest that the clinical benefit would be worth the investment. I appreciate that this is not an economic evaluation, and that your analysis focuses on clinical effectiveness, however, I think that the discussion around the costs needs to be more carefully worded to properly reflect the focus of your study. 4.) Some aspects of your methodology could be clarified, e.g: What is your rationale for selection variables based on significance? Did excluding insignificant variables affect your results? I appreciate that this is common practice in certain strands of the literature, but I have concerns that this can introduce a bias, see, e.g., [1] for an overview. 5.) Why didn’t you pool data from the NDA 2015/16 and 2016/17 for the cross-sectional analyses? This could potentially increase the sample size and precision of the findings. I was also wondering to which extent some of your independent variables where correlated with each other, e.g., since certain agents are systematically associated with weight loss or weight gain 6.) At current, the figures are not very clear and could be improved, e.g., by clearly labeling which section of the figure refers to which outcome. 7.) On a similar notion, on p.2 after Figure 2 there should be a subheading or an introductory sentence to clarify that the next set of results refers to the longitudinal analysis.

	1 Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J Biom Z 2018;60:431–49. doi:10.1002/bimj.201700067

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1
Reviewer Name: Thomas H. Wieringa
Institution and Country: Currently without research affiliation Please state any competing interests or state ‘None declared’: None declared

The topic of your research is interesting and important for improvement of the GP practices. You might consider changing "patient outcomes" as I expected the research investigating patient reported outcomes because of the use of this term.

We appreciate the positive feedback. We have made the change to the term as suggested in the title as elsewhere in the manuscript to ‘Patient Glycaemic Outcome’ – that was a very good point.

In general, the use of the English language is more than sufficient, but there are some little flaws. Please check this.

Duly checked.

If you want your method section to be reproducable, then please elaborate a little more on the statistical analyses (e.g., software used, which longitudinal analyses used, etc.).

Response: This section has been revised accordingly. We have now stated: A consolidated data set was created and extracted using MS Excel 64 bit Power Pivot and extracted data was analysed using the Analyse-it add in, the same multivariate regression methodology as previous published papers

Reviewer: 2
Reviewer Name: Prof. Karel Kostev
Institution and Country: IQVIA, Epidemiology, Germany Please state any competing interests or state ‘None declared’: None declared

Authors present an interesting study. They have developed a statistical model to quantify opportunities for performance improvement for diabetes control within UK GP practices. They used several statistical methods and finally estimated and displayed effects of antihyperglycemic drugs and other factors on the HbA!c values.

Response: Thank you for this feedback. We hope that we have addressed the points below to your satisfaction.

I have the following points.
Major points:

1) Authors have used a multivariate linear regression. For linear regression, the depending variable like HbA1c should have continuous values like , for example, 6.5, 6.6, 6.7 ……. 10.0 and so on.

On the other place they wrote that they defined optimal and poor glycemic control:
“Impact on outcomes was defined by two measures:
% of patients within GPPs at target glycaemic control (TGC) (HbA1c ≤58mmol/mol), and % of patients within GPPs with high glycaemic risk (HGR) (HbA1c >86mmol/mol”
It means that the study patients can reach target or not (yes/no, 1,0). When it looks like that then the use of linear regression would be inappropriate but rather logistic regression should be used.
Instead of ß-values which cannot be understandable interpreted, Odds Ratios could be shown which show very clearly the increasing or decreasing of the probability to achieve the target.

Response: Logistic regression would be appropriate if we had data on each individual patient, but that is not the case. The outcome variable is the proportion of subjects within the practice that have TGC or HGR, and these variables do follow a reasonably normal distributions, so linear regression is appropriate. Odds ratio are not appropriate as we are not considering the probability of an individual falling within a class rather we are considering the change in % of population falling within a class

2) Each antihyperglycemic drug class and drug itself was included as separate group, for example: metformin, DPP-4i, GLP-1 and so on. However, in the diabetes therapy, especially in the case of therapy escalation, most of these drug classes are given as combinations. For example, metformin + DPP-4i, metformin + sulfonylurea, metformin + insulin and so on. The effect of metformin + insulin can be different than the effect of metformin + DPP-4i. GLP-1 + insulin can have another effect as GLP-1+metformin. When the analysis of such combinations is difficult, then authors should mention it in the limitations.

Response: We are evaluating the relative total overall use of different medications within an identified population. The fact that the total of average defined daily dose is actually >1 suggest that patients are either receiving more than one medication or that the actual doses of each medicine being actually prescribed are higher than laid down in the standards. We cannot see at patient level that actual mix of medicines being prescribed and therefore cannot suggest which combination are actually being used. Around 2% of the medication is actually shown as a combination within the BNF and statement ot this effect is now within the report. We have added a statement to this effect

Minor:
- Figures have not an optimal quality and probably other colors should be selected for better presentation. On figures 3 and 4 there is not possible to read the text due to the small font size.
-
Response: The figures have been redrawn with larger font and with use of standardised colours.

Reviewer: 3
Reviewer Name: Peter Eibich
Institution and Country: Max Planck Institute for Demographic Research, Germany Please state any competing interests or state ‘None declared’: None declared.

This study aims to identify factors that affect outcomes for individuals with type 2 diabetes in England. Specifically, the authors focus on the percentage of individuals with type 2 diabetes who achieve target glycaemic control (TGC) and the percentage of patients at higher glycaemic risk (HGR) at the GP practice level. They combine data from several different sources, including the National Diabetes Audit. They estimate cross-sectional and longitudinal (first-differenced) multivariate regression models to examine associations between outcomes (TGC and HGR) and factors such as prescribing, health services and demographics at the practice level. Based on these estimated regression coefficients, the authors then project the change in individuals achieving target glycaemic control or being at higher glycaemic risk if all GP practices would follow best-practice examples. The results indicate that practices using older therapy lines (e.g., sulphonylurea or insulin) show poorer outcomes, whereas

practices using newer agents and a wider range of services (e.g., feet checks) perform better. Following the best-practice could lead to a considerable increase in patients with good outcomes.

This is certainly a very interesting study with a good contribution – identifying practice-styles that are associated with better patient outcomes for individuals with type 2 diabetes. The paper is well-written and concise. However, I have a few concerns, in particular regarding the analysis.

Response

We have employed a relatively novel technique. However this has been validated by peer reviewers of more than 6 recent papers including:
Sodium-glucose co-transporter-2 inhibitors, the latest residents on the block: Impact on glycaemic control at a general practice level in England. Heald AH, Fryer AA, Anderson SG, Livingston M, Lunt M, Davies M, Moreno GYC, Gadsby R, Young RJ, Stedman M. Diabetes Obes Metab 2018; 20: 1659-1669.
Heald AH, Livingston M, Fryer A, Moreno GYC, Malipatil N, Gadsby R, Ollier W, Lunt M, Stedman M, Young RJ. Route to improving Type 1 diabetes mellitus glycaemic outcomes: real-world evidence taken from the National Diabetes Audit. Diabet Med. 2018 Jan; 35: 63-71.

My main concern is that the estimates might be biased by reverse causality. You mention several times throughout the manuscript that there might be confounding, however, you interpret your findings in a causal framework, implying that patient outcomes would improve if prescription would change.

Response: As stated above, we have employed a relatively novel approach here. We report variation between performance in terms of achievement of HbA1c target between thousands of GP practices and separate them by centiles. Our assertions are based around the notion that the reuired underlying behaviours & changes such as training or staff / patient engagement that improve service performance will also underpin the improvement in outcome so that even a proportion of GP practices were brought up to the performance level of the top 10% of GP practices then large numbers of people with T2DM could be brought into target glycaemic control with significant numbers taken out of high glycaemic risk. This not to attribute direct causation, rather to propose that service organisational change and alteration in prescribing practice can have very significant benefits. There is of course a chance that the actual activities carried such as bureacratic or administrative changes are not linked to outcomes

The potential impact of confounders is not discussed (beyond acknowledging that confounding might exist), and you do not conduct any specific sensitivity analyses to address potential confounding. In particular, it seems that you only control for the percentage of patients receiving a certain medication, and not for previous treatments or the treatment line. This could bias your results, since poor outcomes would typically lead to an escalation of treatment, and thus poor outcomes might be systematically associated with different prescription patterns. For example, insulin is typically only prescribed if target glycaemic control cannot be achieved with other therapy lines, i.e., it might not actually be possible to prescribe metformin to patients that previously received insulin (as an example).

Response: Again, if we had access to individual subject data, analysis along these lines would be entirely appropriate, and preferable. However, we only have access to data at the practice level, and the confounding problem will be very different between practices as opposed to between individuals. Certainly, there will be reverse causality in our data: poor outcomes will lead to changes in treatment. However, this should bias the observed effects downwards: if more modern drugs are given to patients who are doing less well, then they will appear to be less effective.

Your finding that, in the cross-sectional analysis, sulphonylurea and insulin are associated with worse outcomes might be an artefact of this reverse causality, since these are typically not prescribed as first-line therapy.

I would suggest to try and distinguish between therapy-lines/therapy patterns if possible, and you will need to discuss the potential bias in your results carefully.

2.) To an extent, this reverse causality bias might be mitigated by the longitudinal analysis, since in this analysis you only relate changes in prescription patterns to changes in your outcomes. However, while you briefly discuss some advantages and disadvantages of the cross-sectional and longitudinal models (mostly regarding the source of the variation), you do not seem to check which of the models is more appropriate (e.g., by using a Hausman test). This makes it difficult for the reader to assess which of the models findings are more reliable. I would suggest to provide some more extensive guidance on the differences in the interpretation of your cross-sectional and longitudinal analyses, and what this might imply for the results.

Response: The cross-sectional model is better at identifying the many factors which are different between different practices but more prone to confounders to the differences between practices. Longitudinal analysis works well for the smaller number of factors where there has been a “net” year on year change and is less prone to confounder. Findings are reinforced when common across both methods/models

3.) In the introduction, you state that “Targeting the more effective agents is therefore fundamental to longer term affordability”. This is not necessarily clear, since the costs of these agents also differ substantially. You acknowledge in the discussion section that using newer drugs would be costly, however, the sentence also seem to suggest that the clinical benefit would be worth the investment. I appreciate that this is not an economic evaluation, and that your analysis focuses on clinical effectiveness, however, I think that the discussion around the costs needs to be more carefully worded to properly reflect the focus of your study.

Response: We accept that we have not included a health economic evaluation. However, this is a separate area that we plan to address in a separate paper. We have altered the discussion section to take account of the point made about careful wording in relation to costs.

4.) Some aspects of your methodology could be clarified, e.g: What is your rationale for selection variables based on significance? Did excluding insignificant variables affect your results? I appreciate that this is common practice in certain strands of the literature, but I have concerns that this can introduce a bias, see, e.g., [1] for an overview.

Response: Selection based on significance can bias effects estimates upwards. However, the magnitude of the bias decreases with the size of the sample, and since we have such a large sample it is likely to be extremely small in this case.

5.) Why didn’t you pool data from the NDA 2015/16 and 2016/17 for the cross-sectional analyses? This could potentially increase the sample size and precision of the findings. I was also wondering to which extent some of your independent variables were correlated with each other, e.g., since certain agents are systematically associated with weight loss or weight gain

Response: We did not pool the data because there were significant changes in prescribing particularly of the SGLT2-is and also of the GLP-1 agonists over this period with year on year reduction in Pioglitzone prescribing – hence combining the data would not be appropriate for these 2 years.

Concerning the second point about how independent variables related to each other

Response: There will be correlations between the predictors, but the coefficients presented are those from the multi-variate model, and hence are controlled for the other variables in the model. They represent the effect of that variable if all other variables are held constant.

6.) At current, the figures are not very clear and could be improved, e.g., by clearly labeling which section of the figure refers to which outcome.

Response: We have done this so that the different sections of the figures are clearer.

7.) On a similar notion, on p.2 after Figure 2 there should be a subheading or an introductory sentence to clarify that the next set of results refers to the longitudinal analysis.

Thanks, we have done this.

1 Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. Biom J Biom Z 2018;60:431–49. doi:10.1002/bimj.201700067

VERSION 2 – REVIEW

REVIEWER	Karel Kostev IQVIA Germany, Epidemiology
REVIEW RETURNED	18-Apr-2019

GENERAL COMMENTS	I thank authors for answers and explatations.

REVIEWER	Peter Eibich Max Planck Institute for Demographic Research, Germany
REVIEW RETURNED	03-May-2019

GENERAL COMMENTS	Most of my comments have been addressed in this revision. In particular, the added discussion around the limitations of the empirical analysis make the manuscript clearer. I have two additional minor comments: 1.) The second sentence in the paragraph titled "Cross-sectional analysis" on p.7 seems to be incomplete. 2.) The figures are still not very easy to read and should be improved before the manuscript can be published. In the copy of the manuscript I have access to, the figures seem to use a grey font in front of a black background, which makes it very difficult to read anything.

VERSION 2 – AUTHOR RESPONSE

Thank you for the comments

We have updated the title to read “Analysis of English General Practice level data linking the levels of demography, service, and specific medication to levels of glycaemic control being achieved in Type 2 Diabetes in order to improve clinical practice and outcomes”.

We have removed the ‘what is known’ and ‘what this paper adds’ sections

Reviewer 3 Comments

The remainder of the sentence of the sentence mentioned seems to be on the next page

Conversion of the figures from provided TIFF by the system to the draft pdf seems to have created this figure formatting issue which is not within our control.