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)	The association of Type 2 diabetes with in-hospital complications among women undergoing breast cancer surgical procedures. A retrospective study using the Spanish National Hospital Discharge Database, 2013–2014.
AUTHORS	Lopez-de-Andres, Ana; Jimenez-Trujillo, Isabel; Hernandez-Barrera, Valentin; de Miguel-Diez, Javier; Mendez-Bailon, Manuel; de Miguel-Yanes, Jose; Perez-Farinos, Napoleon; SALINERO-FORT, MIGUEL; del Barrio, Jose; Romero-Maroto, Martin; Jimenez-Garcia, Rodrigo

VERSION 1 – REVIEW

REVIEWER	David Martinez Clinic Hospital. Madrid Spain
REVIEW RETURNED	14-May-2017

| GENERAL COMMENTS | GENERAL COMMENT
Lopez de Andres et al compare the type of surgical procedures used, comorbidities, in-hospital complications (IHC), and in-hospital outcomes between type 2 diabetes (T2DM) women and age-matched non-diabetic women hospitalized with breast cancer. To do so they use the Spanish National Hospital Discharge Database, 2013–2014.

The authors conclude that women with T2DM who undergo breast cancer surgical procedures have more comorbidity, risk factors and advanced cancer presentation than in similar non- T2DM patients. Mastectomy is more frequently used in diabetic women. Moreover, the procedures in women with T2DM were associated with more IHC. Comorbidity and obesity were strong predictors of IHC in women with T2DM.
This is an interesting well written manuscript. In my opinion can be published after minor revision.

INTRODUCTION
I think for an international reader it would be interesting to know how is the health care system in Spain. Public, private, coverages….

METHODS
The author's state:" If both types of procedures were recorded, that case was excluded." Could this number be added to text?
Please include the number of provinces and the proportion of cases that could be matched. |

Variables included in these models were those with significant results in the bivariate analysis. What p value was considered, <0.05 or 0.1.

RESULTS
In table 4 authors state that only those variable with significant results are shown but age is in the table?. Clarify.

DISCUSSION
Obesity prevalence is very low among diabetic women. Is undercooking possible? Please comment. How could this affect the results?

REVIEWER	gianluca franceschini Catholic University Rome Italy No Competing Interest
REVIEW RETURNED	30-Jun-2017

GENERAL COMMENTS	the authors clearly define their study limitations. In addition, I think there are also some other scientific contents they should explain in order to complete the comparation and to achieve research purpose. I suggest to collect clinical- pathological and anamnestic data, easily reachable in your database, which may influence the evaluation of prognostic outcome such as : - Glicemic status evaluated with any method of diabetes ascertainment (eg. Blood test, self report, blood Hb1AC levels) at the time of breast cancer surgery. - BMI - Type of medical therapy : insulin or oral blood glucose lowering drugs. - Type and duration of antibiotic prophylaxis for the breast surgery - Number of mastectomies in diabetes patients undergone to a prosthetic reconstruction and the eventual use of an acellular dermal matrix."

REVIEWER	Iliana Lega Women's College Research Institute Toronto, Canada
REVIEW RETURNED	03-Aug-2017

GENERAL COMMENTS	This large, population-based study uses Spanish databases to examine the difference in rate of surgical procedures (mastectomy vs BCS), as well as in-hospital complications and outcomes between women with Bca with and without diabetes. Women with DM and Bca who underwent surgery were matched on age and area of residence to women without diabetes, matching was not done on stage. Though many covariates were included based on data from discharge databases, the validity of these data is unclear. Overall, this study is clinically very relevant and important for betting understanding treatment and outcome differences among breast cancer patients with and without diabetes. A better description of the validity of the databases used and the minor changes to the analysis would strengthen this study.

Covariates:

Since the authors are relying so heavily on these hospital discharge databases for information on all covariates and the exposure of interest, more detail and attention needs to be given to how these databases are organized, how accurate the information is, how far is the lookback for identifying these conditions, how the information is entered etc. If there has been validation of these databases, those studies should be at least referenced here as well. For the diabetes variable in particular, the authors should discuss in more detail how a diagnosis of DM is made to qualify for entry in the discharge database (i.e is it based on medication use? Glycemic control/HbA1c? Previous visits/documentation/history of DM? in-hospital diagnosis? etc) and whether this method of diagnosing diabetes has been validated. Since the discharge summary may also contain new diagnoses (I assume), how do the authors know that this is not an incident diagnosis of diabetes made during the hospital stay? For other covariates like 'acute myocardial infarction' the authors need to specify if this is a condition that occurred during the hospitalization or is part of the patient's past medical history. Lastly, the authors need to define 'obesity' (ie. what is the BMI cutoff? Is it obesity at time of diagnosis or based on historical data?) and again give information on how this covariate is ascertained. In their discussion they state that 'obesity may not be adequately reported in the database'. First of all, this should be stated up front instead of only in the limitations section. Second, the authors needs to minimize the discussion surrounding obesity in their text and conclusions if they are not sure on the validity of this variable to begin with.

A thorough description of the data should be included in the methods, and the limitations clearly stated in the discussion.

Analysis: Since the authors have information on stage, I would consider matching the diabetes and non-diabetes group on this variable as well. Stage is likely a confounder (though the results found by the authors were not significant) in their analysis since it can impact intensity of treatments received and may reflect on the overall wellness of the patient. Furthermore, matching on stage would be important for comparing surgical procedures as there may be different surgical indications based on tumour size/stage.

In the logistic regression, I think it would be more clinically meaningful to analyze CCI in the 3 categories previously used in table 1-3 instead of as a mean. No comorbidities would be the referent group and you could determine the impact of increasing comorbidities on the outcomes.

For the final analysis seeking to identify variables that predict in hospital complications I would list all the variables included in the model in the methods section (Pg 9) and include estimates for both significant and non-significant covariates in table 4.

Minor:

All tables should be at the end of the text.

Table 1-3 are a little redundant and make the manuscript very busy. Perhaps all this information could be put into one table, the first 2 columns would be for the whole cohort, columns 3-4 for the BCS, and the next two columns for mastectomy?

pg 12, ln 10 – the differences in sentinel node dissection should be followed by a p value

In general, the terminology 'patients with diabetes' and 'patients without diabetes' is preferred to diabetic/non-diabetic patients.

Table 4: The subgroups (Mastectomy and BCS) can be listed as columns after the results for the entire group, instead of rows underneath. It would make the table easier to read.

Reviewer: 1
Reviewer Name: David Martinez
Institution and Country: Clinic Hospital, Madrid, Spain Competing Interests: None declared

GENERAL COMMENT
Lopez de Andres et al compare the type of surgical procedures used, comorbidities, in-hospital complications (IHC), and in-hospital outcomes between type 2 diabetes (T2DM) women and age-matched non-diabetic women hospitalized with breast cancer. To do so they use the Spanish National Hospital Discharge Database, 2013–2014.
The authors conclude that women with T2DM who undergo breast cancer surgical procedures have more comorbidity, risk factors and advanced cancer presentation than in similar non- T2DM patients. Mastectomy is more frequently used in diabetic women. Moreover, the procedures in women with T2DM were associated with more IHC. Comorbidity and obesity were strong predictors of IHC in women with T2DM.
This is an interesting well written manuscript. In my opinion can be published after minor revision.

ANSWER
Thank you for your constructive comments that will surely improve the manuscript.

COMMENT
INTRODUCTION
I think for an international reader it would be interesting to know how is the health care system in Spain. Public, private, coverages….

ANSWER
The following sentence has been added to the introduction
"The Spanish Health Care System is public and offers universal coverage with no out-of-pocket expenses for patients."

COMMENT
METHODS
The author's state:" If both types of procedures were recorded, that case was excluded." Could this number be added to text?

ANSWER
The number of patients with both types of procedures recorded and excluded was 137. This number has been added to the text.

COMMENT
Please include the number of provinces and the proportion of cases that could be matched.
ANSWER
Spain has 50 provinces. This number has been added to the text. The proportion of cases matched was 98.6% when we matched by age and province of residence and 88% when we matched by age, province of residence, type of procedure and stage.
These proportions have been added to the methods section.

COMMENT
Variables included in these models were those with significant results in the bivariate analysis. What p value was considered, <0.05 or 0.1.

ANSWER
We considered p<0.05. This value has been added to the text.

COMMENT
RESULTS
In table 4 authors state that only those variable with significant results are shown but age is in the table?. Clarify.

ANSWER
Sorry for this mistake. The table shows the results of all variables included in the model, significant and not significant. The footnote of the table has been deleted to clarify this point.

COMMENT
DISCUSSION
Obesity prevalence is very low among diabetic women. Is undercooking possible? Please comment. How could this affect the results?

ANSWER
You are right. The following paragraph, with four new references, has been added to the discussion section:
"However, in our study population, the prevalence of obesity is very low among women with diabetes, which is possibly a consequence of under-recording this condition. Previous studies conducted in Spain and other countries have also found an under-reporting of obesity in administrative data. [20-23] These authors suggest the following possible reasons to explain the low rates of obesity in administrative data: i) obesity is not explicitly mentioned in physician reports; ii) people who codify may not record obesity owing to time constraints when performing data abstraction; iii) when time for coding is limited, coders tend to include severe conditions but not risk factors and; iv) the diagnosis of obesity is often not based on the BMI, but rather on the subjective observation made by the clinician, which means that more severe obesity is over-codified [21-23] Furthermore, as described earlier, a possible differential information bias may occur and thus the misclassification of obesity may be related to the presence of diabetes. Thus, ICD codes for obesity may be more commonly assigned to patients suffering from other comorbidities (including diabetes) or postoperative complications. This suggests a greater association between obesity and adverse events than what is obtained based on the BMI calculations. [21, 23] As a result, any association between obesity and the presence of in-hospital complications must be interpreted with caution."
20. Sánchez-Muñoz LA. Nutritional status, heart failure and minimum basic data set. Rev Esp Cardiol (Engl Ed). 2012 Jun;65(6):583.
21. Martin BJ, Chen G, Graham M, et al. Coding of obesity in administrative hospital discharge abstract data: accuracy and impact for future research studies. BMC Health Serv Res. 2014 Feb 13;14:70.
22. Lujic S, Watson DE, Randall DA, et al. Variation in the recording of common health conditions in routine hospital data: study using linked survey and administrative data in New South Wales, Australia. BMJ Open. 2014 Sep 3;4(9):e005768.
23. McLynn RP, Geddes BJ, Cui JJ, et al. Inaccuracies in ICD Coding for Obesity Would be Expected to Bias Administrative Database Spine Studies Toward Overestimating the Impact of Obesity on Perioperative Adverse Outcomes. Spine (Phila Pa 1976). 2017 Aug 1. doi: 10.1097/BRS.0000000000002356.

Reviewer: 2
Reviewer Name: Gianluca Franceschini
Institution and Country: Catholic University Rome, Italy Competing Interests: no

COMMENT
The authors clearly define their study limitations. In addition, I think there are also some other scientific contents they should explain in order to complete the comparation and to achieve research purpose.
I suggest to collect clinical- pathological and anamnestic data, easily reachable in your database, which may influence the evaluation of prognostic outcome such as:
Glicemic status evaluated with any method of diabetes ascertainment (eg. Blood test, self report, blood Hb1AC levels) at the time of breast cancer surgery.
- BMI
- Type of medical therapy : insulin or oral blood glucose lowering drugs.
- Type and duration of antibiotic prophylaxis for the breast surgery
- Number of mastectomies in diabetes patients undergone to a prosthetic reconstruction and the eventual use of an acellular dermal matrix."

ANSWER
Thank you very much for your comments. Unfortunately information on laboratory results, BMI, medical treatments (glucose lowering drugs or antibiotics), patients who underwent a prosthetic reconstruction and the eventual use of an acellular dermal matrix are not included in the Spanish National Hospital Discharge Database. This information has been added to the limitations section.

Reviewer: 3
Reviewer Name: Iliana Lega
Institution and Country: Women's College Research Institute, Toronto, Canada Competing Interests: None declared

COMMENT
This large, population-based study uses Spanish databases to examine the difference in rate of surgical procedures (mastectomy vs BCS), as well as in-hospital complications and outcomes between women with Bca with and without diabetes. Women with DM and Bca who underwent surgery were matched on age and area of residence to women without diabetes, matching was not done on stage.

ANSWER
We agree that matching by stage is very relevant so we have also matched by this variable. The text and tables have been changed in all sections as a consequence of this new matching variable.

COMMENT
Though many covariates were included based on data from discharge databases, the validity of these data is unclear. Overall, this study is clinically very relevant and important for betting understanding treatment and outcome differences among breast cancer patients with and without diabetes. A better description of the validity of the databases used and the minor changes to the analysis would strengthen this study.

Covariates:
Since the authors are relying so heavily on these hospital discharge databases for information on all covariates and the exposure of interest, more detail and attention needs to be given to how these databases are organized, how accurate the information is, how far is the lookback for identifying these conditions, how the information is entered etc. If there has been validation of these databases, those studies should be at least referenced here as well.

ANSWER
The following paragraph, with five new references, has been added to the methods section regarding the characteristics of the Spanish National Hospital Discharge Database.
"The SNHDD was implemented in 1987. According to Spanish legislation, all public and private Spanish hospitals must periodically submit data to the health authorities regarding those patients who have been hospitalized for at least 24 hours. The information required includes patient's characteristics such as their clinical history number, personal ID number, date of birth, sex, country of birth and address. Clinical variables include: admission and discharge dates, up to 14 diagnoses at discharge and up to 20 procedures performed during the hospital stay. The primary/main diagnosis is defined as the condition which, after proper investigation, is considered the reason why the patient was admitted to the hospital. The secondary diagnosis includes those diseases or risk factors that coexist with the primary diagnosis at the time of admission or were detected during the hospitalization and that, in the opinion of the treating physician, may have affected the patient's progress or treatment plan.

Procedures include those diagnostic or therapeutic procedures conducted during hospitalization. Information on the service where the patient has received care and the type of discharge (home, decease, voluntary discharge, other hospital, and social institution) is also collected.
All patients discharged from the hospital must have a discharge report signed by the physician discharging the patient that includes the above noted information. The Codification Unit of the hospital uses this discharge report and any additional information required by the hospital databases to complete the SNHDD. The database uses the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) for coding.
The Spanish Ministry of Health conducts regular audits to assess the accuracy of the SNHDD [8, 9]. Several studies have been conducted to assess the validity of the SNHDD for several conditions, including diabetes. [10-13]"
9. Estudio de las comorbilidades que componen el índice de Elixhauser. Análisis de prevalencia y fiabilidad en los registros del CMBD estatal de hospitalizacion Available from http://icmbd.es/docs/informe-comorbilidades-1.pdf Accessed 30 August 2017.
10. Ribera A, Marsal JR, Ferreira-González I, Cascant P, et al. Predicting in-hospital mortality with coronary bypass surgery using hospital discharge data: comparison with a prospective observational study. Rev Esp Cardiol 2008; 61(8):843-852.
11. Rodrigo-Rincón I, Martin-Vizcaíno MP, Tirapu-León B, et al. Usefulness of administrative databases for risk adjustment of adverse events in surgical patients. Cir Esp. 2016;94(3):165-74.
12. Calle JE, Saturno PJ, Parra P, et al. Quality of the information contained in the minimum basic data set: Results from an evaluation in eight hospitals. Eur J Epidemiol. 2000; 16: 1073-80
13. Hernández Medrano I, Guillán M, Masjuan J, et al. Reliability of the minimum basic dataset for diagnoses of cerebrovascular disease. Neurologia. 2017;32(2):74-80.

COMMENT
For the diabetes variable in particular, the authors should discuss in more detail how a diagnosis of DM is made to qualify for entry in the discharge database (i.e is it based on medication use? Glycemic control/HbA1c? Previous visits/documentation/history of DM? in-hospital diagnosis? etc) and whether this method of diagnosing diabetes has been validated.

ANSWER

The following paragraphs have been added to the text:

"In Spain, the diagnosis of type 2 diabetes is mainly performed in primary care centers. [35] Even if patients are diagnosed during hospitalization they are sent to their primary care center for the follow up and necessary medical prescriptions. A study conducted to validate the diabetes mellitus diagnosis in the computerized clinical records of primary health care, taking the 2003 American Diabetes Association Consensus Statement as the gold standard, found that the agreement was very high (κ=0.990), with a specificity of 99.49% and a sensitivity of 99.53%. [35] The validity of the diabetes diagnosis in the SNHDD has been assessed in two previous studies, revealing a sensitivity of 55% and 63.7% and a specificity of approximately 97%.[10,11]

The only moderate sensitivity found means that an important proportion of T2DM patients do not have this diagnosis codified in their discharge report. On the other hand, the very high specificity means that most patients without a T2DM diagnosis do not really have this disease; therefore, we think that the effect of this misclassification on our design is possibly very small."

35. de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, et al. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. BMC Med Res Methodol. 2011;11:146.

COMMENT

Since the discharge summary may also contain new diagnoses (I assume), how do the authors know that this is not an incident diagnosis of diabetes made during the hospital stay? For other covariates like 'acute myocardial infarction' the authors need to specify if this is a condition that occurred during the hospitalization or is part of the patient's past medical history.

ANSWER

You are right. The following paragraph has been added to the limitations section. "….unfortunately the ICD 9 does not include information regarding whether a condition is part of the patient's past medical history or if this appeared during the hospitalization. Therefore, it is possible that a patient may first be diagnosed with diabetes or any other condition studied during the hospitalization. However, we think that this possibly affects a very small proportion of patients and thus the effect on our results would be minimal."

COMMENT

Lastly, the authors need to define 'obesity' (ie. what is the BMI cutoff? Is it obesity at time of diagnosis or based on historical data?) and again give information on how this covariate is ascertained. In their discussion they state that 'obesity may not be adequately reported in the database'. First of all, this should be stated up front instead of only in the limitations section. Second, the authors needs to minimize the discussion surrounding obesity in their text and conclusions if they are not sure on the validity of this variable to begin with.

A thorough description of the data should be included in the methods, and the limitations clearly stated in the discussion.

ANSWER

Thank you for this comment. You are right. The text and the conclusion regarding the role of obesity have been minimized. The following paragraph, with four new references, has been added to the discussion regarding the limited validity of obesity in administrative data.

"However, in our study population, the prevalence of obesity is very low among women with diabetes, which is possibly a consequence of under-recording this condition.

Previous studies conducted in Spain and other countries have also found an under-reporting of obesity in administrative data. [20-23] These authors suggest the following possible reasons to explain the low rates of obesity in administrative data: i) obesity is not explicitly mentioned in physician reports; ii) people who codify may not record obesity owing to time constraints when performing data abstraction; iii) when time for coding is limited, coders tend to include severe conditions but not risk factors and; iv) the diagnosis of obesity is often not based on the BMI, but rather on the subjective observation made by the clinician, which means that more severe obesity is over-codified [21-23] Furthermore, as described earlier, a possible differential information bias may occur and thus the misclassification of obesity may be related to the presence of diabetes. Thus, ICD codes for obesity may be more commonly assigned to patients suffering from other comorbidities (including diabetes) or postoperative complications. This suggests a greater association between obesity and adverse events than what is obtained based on the BMI calculations. [21, 23] As a result, any association between obesity and the presence of in-hospital complications must be interpreted with caution."

20. Sánchez-Muñoz LA. Nutritional status, heart failure and minimum basic data set. Rev Esp Cardiol (Engl Ed). 2012 Jun;65(6):583.
21. Martin BJ, Chen G, Graham M, et al. Coding of obesity in administrative hospital discharge abstract data: accuracy and impact for future research studies. BMC Health Serv Res. 2014 Feb 13;14:70.
22. Lujic S, Watson DE, Randall DA, et al. Variation in the recording of common health conditions in routine hospital data: study using linked survey and administrative data in New South Wales, Australia. BMJ Open. 2014 Sep 3;4(9):e005768.
23. McLynn RP, Geddes BJ, Cui JJ, et al. Inaccuracies in ICD Coding for Obesity Would be Expected to Bias Administrative Database Spine Studies Toward Overestimating the Impact of Obesity on Perioperative Adverse Outcomes. Spine (Phila Pa 1976). 2017 Aug 1. doi: 10.1097/BRS.0000000000002356.

COMMENT
Analysis: Since the authors have information on stage, I would consider matching the diabetes and non-diabetes group on this variable as well. Stage is likely a confounder (though the results found by the authors were not significant) in their analysis since it can impact intensity of treatments received and may reflect on the overall wellness of the patient. Furthermore, matching on stage would be important for comparing surgical procedures as there may be different surgical indications based on tumour size/stage.

ANSWER
We agree that matching by stage is very relevant so we have also matched by this variable. The text and tables have been changed in all the sections as a consequence of this new matching variable.

COMMENT
In the logistic regression, I think it would be more clinically meaningful to analyze CCI in the 3 categories previously used in table 1-3 instead of as a mean. No comorbidities would be the referent group and you could determine the impact of increasing comorbidities on the outcomes
For the final analysis seeking to identify variables that predict in hospital complications I would list all the variables included in the model in the methods section (Pg 9) and include estimates for both significant and non-significant covariates in table 4.

ANSWER
Following your suggestion in the logistic regression model we have analyzed CCI in the 3 categories, we have listed all the variables included in the model in the methods section and included estimates for both significant and non-significant covariates in table 4 (now Table 2)

COMMENT
Minor
All tables should be at the end of the text.

ANSWER
Following your suggestion all tables have been moved to the end of the text.

COMMENT
Table 1-3 are a little redundant and make the manuscript very busy. Perhaps all this information could be put into one table, the first 2 columns would be for the whole cohort, columns 3-4 for the BCS, and the next two columns for mastectomy?

ANSWER
Following your suggestion Tables 1-3 have been joined in a single table.

COMMENT
pg 12, ln 10 – the differences in sentinel node dissection should be followed by a p value.

ANSWER
The p value has been added to the differences in sentinel node dissection.

COMMENT
In general, the terminology 'patients with diabetes' and 'patients without diabetes' is preferred to diabetic/non-diabetic patients.

ANSWER
Following your suggestion "diabetic/non-diabetic patients" has been replaced by 'patients with diabetes' and 'patients without diabetes' in the text and tables.

COMMENT
Table 4: The subgroups (Mastectomy and BCS) can be listed as columns after the results for the entire group, instead of rows underneath. It would make the table easier to read.

ANSWER
Following your suggestion the subgroups (Mastectomy and BCS) are listed as columns after the results for the entire group in Table 2.

VERSION 2 – REVIEW

REVIEWER	David Martinez Complutense University, Spain
REVIEW RETURNED	08-Sep-2017

GENERAL COMMENTS	My comments have been answered correctly

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: David Martinez
Institution and Country: Complutense University, Spain
Competing Interests: None declared

COMMENT
My comments have been answered correctly

ANSWER
Thank you very much