PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	HERPES ZOSTER RELATED HEALTHCARE BURDEN AND COSTS IN IMMUNOCOMPROMISED (IC) AND IC-FREE POPULATIONS IN ENGLAND: AN OBSERVATIONAL RETROSPECTIVE DATABASE ANALYSIS
AUTHORS	Curran, Desmond; Hunjan, Manjit; El Ghachi, Amale; El-Hahi, Yassine; Bianco, Veronique; Ferreira, Germano

VERSION 1 - REVIEW

REVIEWER	John Sampalis McGill university, Canada
REVIEW RETURNED	13-May-2018

GENERAL COMMENTS	This is excellent work. However, I have some minor concerns about some of the exclusion criteria. Specifically, the exclusion of patients without date of diagnosis of the IC condition and the exclusion of patients with PHN > 365 days. It would be useful to have some description of the observations excluded and the potential impact on the validity of the results. My most serious concern is with the analyses conducted and the presentation of the results. We are only being provided with means and no indication of the variance such as SD SEM or Range. This is very concerning. We all know that HCRU and Cost estimates are not normally distributed and hence the mean alone may be meaningless and bias. I would like to see more details on the distribution of HCRU including the median value and area plots to show the distribution of HCRU and costs for each group. Accepting that given the sample size all t-tests would be statistically significant, presenting 99% confidence interval's of the difference would be useful for the reader to assess the results. There is another concern with the analyses. This is related to the fact that HCRU and related costs are a function of follow up duration. In fact the longer the follow up the higher the HCRU will be. Hence follow up duration is part of the outcome. This is why Poisson distributions are best suited for this analyses since they do take into account the duration of follow up. At the very least the authors must report HCRU in terms of Incidence Density Rates (Events per 100 person months of follow up, as an example) and between group differences must be assess with the Incidence Density Rate Ratio (IDRR) using 95% or 99% confidence intervals of the IDRR to

	assess precision and statistical significance. Finally although the cohorts are matched for demographics, there may be other factors that contribute to increased HCRU and Health care costs in the IC group and this may be related to HZ indirectly. Using multivariate models to control for some of these may be helpful, but not critical. Another comment has to do with the perception of bias or conflict of interest in the study. Although there is no reason to doubt the integrity and credibility of the authors, the fact that this study has been conducted entirely by a company that has significant investments in vaccines may raise some concerns. I would recommend that the authors consider the involvement of an academic researcher that would participate in the analysis and or interpretation of the study results. Related to this perception of conflict of interest, the statement in the discussion regarding the new vaccines that are being developed, appears promotional, and must be removed.

REVIEWER	Harriet Forbes LSHTM, UK
REVIEW RETURNED	08-Oct-2018

GENERAL COMMENTS	This is a well-written paper which adds to the existing literature on zoster-related healthcare costs. Some minor points which could be clarified: - In the final paragraph of the introduction, some results are included. If these relate to the data in the manuscript, please move to the results section. Or if they relate to previous work, move it to an earlier section of the introduction. - The study follow-up ended in March 2012. Was this for any particular reason? - Why did you look for records 7 days prior to zoster diagnosis date? This isn't necessarily wrong, but needs explanation. - Can you be clearer throughout the paper (both in the text and tables) that you looked for HZ-related records, rather than any records, in the -7 to 365 days from zoster onset? - Can you signpost the reader to where they can find the Read/ICD-10/BNF codes throughout the methods section? - At the start of the discussion it would be helpful to include your actual cost estimates.

REVIEWER	Qian Li Evidera, US
REVIEW RETURNED	24-Oct-2018

GENERAL COMMENTS	This study assessed the healthcare resource utilization (HCRU) and costs of herpes zoster (HZ), with or without postherpetic neuralgia (PHN), among patients with immunocompromised (IC) conditions from large electronic health record databases in the UK. IC-free individual, who were matched to the IC patients by age, gender and practice location, were used as a control group. The findings were consistent with previous studies: patients with IC conditions incur higher HZ-related healthcare costs than IC-free individuals. Below is a list of my major comments:

	General question on method. Although the outcomes are HZ-related HCRU and costs, it will be helpful to report some baseline values of all-cause HCRU and costs since these reflect patient's general health status and healthcare seeking behaviors. The IC and IC-free individuals were not matched on any HCRU and costs, therefore knowing how HCRU and costs were different at baseline between the two cohorts will help to further evaluate the findings. General question on data source. Data source is from year 2000 to 2012. Why don't use more recent data? General question on results. How can the findings be generalizable? Since the study population is from UK, can the results be generalized to national level? METHODS, Row 139-141: It is not clear when the follow-up started (since IC diagnosis or HZ diagnosis?). And what is the time window to define IC population (with IC diagnosis during 12 months after HZ diagnosis?). METHODS, Row 152-153: Please clarify how the patients were selected. My understanding is that HZ and IC patients were matched to HZ and IC-free patients, so the sample selection flow should be 1) selecting patients with HZ, 2) identifying IC and IC-free cohorts, and 3) matching IC-free cohort to IC cohort. But the paper reads like 2)□3) □1). DISCUSSION Row 248-251: The paragraph mentioned "the calculation of IC condition prevalence rates, HZ incidence rates". But these outcomes were not assessed at all in the paper. Please clarify. DISCUSSION Row 283-285: This paragraph mentioned "higher incidence rates of HZ only". Please provide a reference for this. DISCUSSION Row 287-298: This paragraph only cited the previous published study. The authors could elaborate how their study different from the previous studies and their contribution to the literature. Since the methods used in this study is quite straight forward, compared to some previous studies, the authors could use discuss the strength (large sample size?) of their paper here. CONCLUSION Row 308-309. The conclusion stated that "In conclusion, individuals with IC conditions incurred higher healthcare utilization and costs than IC-free individuals." Please put more specifications on this conclusion, for example, in UK, patients seeking healthcare in general practices.

VERSION 1 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 1

Reviewer Name: John Sampalis

This is excellent work.

Thank you for your kind words.

However, I have some minor concerns about some of the exclusion criteria. Specifically, the exclusion of patients without date of diagnosis of the IC condition and the exclusion of patients with PHN > 365 days. It would be useful to have some description of the observations excluded and the potential impact on the validity of the results.

Patients without a date of diagnosis of the IC condition were excluded, as without a diagnosis date, it is not possible to determine if the HZ episode occurred before or after the diagnosis date.

The sentence “Individuals who recorded the first PHN event date after 365 days post HZ event date were excluded.” was actually incorrect/misleading as only the data occurring after 365 days post HZ event date was excluded. Thus, the sentence has been reworded “Consequently, individuals who recorded the first PHN event date after 365 days post HZ event date were classified as not having PHN.” Note, it is possible that these were also recurrent episodes, so one can consider that this was a conservative approach.

My most serious concern is with the analyses conducted and the presentation of the results. We are only being provided with means and no indication of the variance such as SD SEM or Range. This is very concerning. We all know that HCRU and Cost estimates are not normally distributed and hence the mean alone may be meaningless and bias. I would like to see more details on the distribution of HCRU including the median value and area plots to show the distribution of HCRU and costs for each group. Accepting that given the sample size all t-tests would be statistically significant, presenting 99% confidence interval's of the difference would be useful for the reader to assess the results.

As the reviewer correctly pointed out HCRU and costs are not normally distributed as such we’ve added the medians and standard deviations to the cost tables in the manuscript (i.e. Tables 2 and 3). Adding more detail could make the tables unreadable. We’ve also added text to the results to highlight this concern “It is also noteworthy that the means are consistently higher than medians, and as is common for healthcare cost data, the distribution is skewed to the right.”. We believe that for an individual the “typical value” may be better represented by the median, however, for a population we

believe that the mean will be more meaningful as there will always be outliers and from a budget/cost point of view, from a population perspective, the means are more relevant as a summary statistic.

There is another concern with the analyses. This is related to the fact that HCRU and related costs are a function of follow up duration. In fact the longer the follow up the higher the HCRU will be. Hence follow up duration is part of the outcome. This is why Poisson distributions are best suited for this analyses since they do take into account the duration of follow up. At the very least the authors must report HCRU in terms of Incidence Density Rates (Events per 100 person months of follow up, as an example) and between group differences must be assess with the Incidence Density Rate Ratio (IDRR) using 95% or 99% confidence intervals of the IDRR to assess precision and statistical significance.

The reviewer is correct that HCRU and costs are a function of follow-up time. In our previous publication we have reported the incidence of HZ cases, i.e. of HZ per 1000 person-years (PY) calculated as the total number of incident HZ cases divided by the sum of the number of PY at risk (Yanni et al. 2018). In this analysis we analysed the costs per HZ episode (which is not dependent on the total follow-up time) and also over fixed periods of follow-up time per HZ episode. Costs of HZ only cases were assessed during the period 7 days prior to 30 days post HZ onset. The costs of PHN were analyzed over 2 time-periods, i.e. (1) 7 days prior to 90 days post HZ onset and (2) 7 days pri-or to 365 days post HZ onset. Note, the most frequently used definition of PHN is: pain persisting or appearing at least 90 days following rash onset. The rationale for the time periods studied for PHN was that using analysis period 1 alone could lead to an underestimation of PHN costs where-as using analysis period 2 only could overestimate these costs.

Finally although the cohorts are matched for demographics, there may be other factors that contribute to increased HCRU and Health care costs in the IC group and this may be related to HZ indirectly. Using multivariate models to control for some of these may be helpful, but not critical.

We agree with the reviewer that there may be other factors that contribute to increased HCRU and Health care costs. We performed these analyses as descriptive analysis only. We are interested in describing the HZ-related costs in the 2 populations (i.e. IC and IC-Free). Our aim was not to demonstrate that there are significant differences between them.

Another comment has to do with the perception of bias or conflict of interest in the study. Although there is no reason to doubt the integrity and credibility of the authors, the fact that this study has been conducted entirely by a company that has significant investments in vaccines may raise some concerns. I would recommend that the authors consider the involvement of an academic researcher that would participate in the analysis and or interpretation of the study results. Related to this perception of conflict of interest, the statement in the discussion regarding the new vaccines that are being developed, appears promotional, and must be removed.

We agree with the reviewer that there may be a perception of bias regarding the results given that all authors are associated with a pharmaceutical company. We believe that our results are however

unbiased and are consistent with the findings of previous studies as noted by reviewer 3. We also noted in the discussion that our results are consistent with a number of independently run studies. As the study is already completed we believe it is too late to consider the involvement of an academic researcher. Based on the reviewer's comments we have deleted the text in the conclusion regarding vaccines (both new and old).

The protocol was approved by Independent Scientific Advisory Committee (ISAC) before the analysis was performed, i.e. ISAC reviewed and approved the scientific content of the protocol. This is described in the methods section "The study protocol was approved by the Independent Scientific Advisory Committee (ISAC) for the Medicines and Healthcare Products Regulatory Agency database research (ISAC protocol number 14_222R)." We have also added a few standard sentences in the acknowledgements section, i.e. "This study is based in part on data from the Clinical Practice Research Datalink obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. However, the interpretation and conclusions contained in this report are those of the authors alone." We've completed the analysis as specified in the protocol. We would however need to go back to ISAC to get approval to perform additional analyses.

Reviewer: 2

Reviewer Name: Harriet Forbes

This is a well-written paper which adds to the existing literature on zoster-related healthcare costs.

Thank you for your kind words.

Some minor points which could be clarified:

- In the final paragraph of the introduction, some results are included. If these relate to the data in the manuscript, please move to the results section. Or if they relate to previous work, move it to an earlier section of the introduction.

These results are from our previous publication. We have edited the wording to make it clearer: "The clinical burden of disease epidemiological results of the study are reported elsewhere7, and may be summarized as follows: the prevalence of IC conditions increased from 7.6% in individuals aged 18-44 YOA to 42.2% in individuals aged ≥80 YOA; the incidence rate of HZ in the IC cohort was 3.5/1000 PY in individuals aged 18-49 YOA increasing to 12.6/1000 PY in individuals aged ≥80 YOA."

- The study follow-up ended in March 2012. Was this for any particular reason?

As this was quite a complex study it took some time to run. Initially, we began developing the study protocol in 2014 and received Independent Scientific Advisory Committee (ISAC) approval to use CPRD and HES data up until March 2012. The analysis began in 2015 and the final study report was completed in 2016. The publication of the epidemiological results, which contains much detail on the study design, was accepted for publication in BMJ open earlier this year (Yanni et al. 2018). We decided to wait until the Epi paper was accepted before submitting the HCRU and cost paper.

- Why did you look for records 7 days prior to zoster diagnosis date? This isn't necessarily wrong, but needs explanation.

Many studies on HCRU and costs include a number of days prior to diagnosis as their may be a delay in diagnosis and HCRU may be utilized prior to diagnosis. For example, Yawn et al. 2009, included 14 days prior to diagnosis and Li et at. 2016 (i.e. reviewer 3) included 21 days prior to diagnosis in their analysis of healthcare costs attributable to HZ. We added a sentence to the discussion, i.e. “Many studies on HCRU and costs include a number of days prior to diagnosis, e.g. 14 or 21 days, as their may be a delay in diagnosis and HCRU may be utilized prior to diagnosis [Yawn et al. 2009, Li et at. 2016].”

- Can you be clearer throughout the paper (both in the text and tables) that you looked for HZ-related records, rather than any records, in the -7 to 365 days from zoster onset?

We added a few sentences to the discussion, i.e. “In this study, every effort was made to include only resources directly related to HZ. For example, only hospitalized patients were included with an ICD-10 HZ diagnosis identified in the HES database. Similarly, only medications potentially related to HZ treatment were included (see Supplementary Text Tables 2 and 4).”

- Can you signpost the reader to where they can find the Read/ICD-10/BNF codes throughout the methods section?

We added a sentence at the end of the methods section as follows “Further details, including information on the IC populations included, ICD-10 codes for HZ and PHN, and unit healthcare costs are provided in the Supplementary Text.”

- At the start of the discussion it would be helpful to include your actual cost estimates.

We added a sentence to the discussion, i.e. “HZ-related mean treatment costs per patient were higher in IC individuals (£189 versus £104 in IC and IC-free individuals aged 18-49 YOA, respectively increasing to £557 versus £401 in IC and IC-free individuals aged ≥80 YOA, respectively).”

Reviewer: 3

Reviewer Name: Qian Li

This study assessed the healthcare resource utilization (HCRU) and costs of herpes zoster (HZ), with or without postherpetic neuralgia (PHN), among patients with immunocompromised (IC) conditions from large electronic health record databases in the UK. IC-free individual, who were matched to the IC patients by age, gender and practice location, were used as a control group.

We added in the discussion “In this study, we presented the healthcare resource utilization and costs associated with HZ in both IC and IC-free populations using large electronic health record databases in the UK.”

The findings were consistent with previous studies: patients with IC conditions incur higher HZ related healthcare costs than IC-free individuals.

Thank you for this affirmation of the study results.

Below is a list of my major comments:

General question on method. Although the outcomes are HZ-related HCRU and costs, it will be helpful to report some baseline values of all-cause HCRU and costs since these reflect patient’s general health status and healthcare seeking behaviors. The IC and IC-free individuals were not matched on any HCRU and costs, therefore knowing how HCRU and costs were different at baseline between the two cohorts will help to further evaluate the findings.

As the reviewer suggested HCRU will differ between the IC and the IC-free individuals. However, if we matched patients on their HCRU then we would potentially match the sickest of the IC-free patients with the healthiest of the IC patients. It was not our aim to estimate the incremental costs or to demonstrate that there are significant differences between the HZ costs in the 2 populations (i.e. IC and IC-Free). Also, we made effort to include only healthcare costs directly related to HZ. We added a few sentences to the discussion, i.e. “In this study, every effort was made to include only resources directly related to HZ. For example, only hospitalized patients were included with an ICD-10 HZ diagnosis identified in the HES database. Similarly, only medications potentially related to HZ treatment were included (see Supplementary Text Tables 2 and 4).”

To illustrate how general HCRU (i.e. non-HZ related) were different between the two cohorts we added a table to the supplementary text, i.e. Supplementary Table 6. We also added the following text in the results section “Supplementary Table 6 presents the non-HZ related hospital inpatient stay for the period 7 days to 365 days post initial-HZ onset. The mean number of non-HZ related

hospitalizations were consistently higher in IC patients compared to and IC-free patients and increased with age."

General question on data source. Data source is from year 2000 to 2012. Why don't use more recent data?

See comment to reviewer 2

General question on results. How can the findings be generalizable? Since the study population is from UK, can the results be generalized to national level?

Herrett et al. 2015 reported that approximately 6.9% of the UK population are included in the CPRD database and patients are broadly representative of the UK general population in terms of age, sex and ethnicity.

METHODS, Row 139-141: It is not clear when the follow-up started (since IC diagnosis or HZ diagnosis?). And what is the time window to define IC population (with IC diagnosis during 12 months after HZ diagnosis?).

Follow-up started in 2002. Further details are provided in the supplementary text as follows "HZ cases identified in the Clinical Practice Research Datalink (CPRD) database were defined as subjects reporting at least one HZ-related READ code. Incident cases were subjects with at least 12 months of active registration in CPRD and no past record of HZ diagnosis during at least 12 months prior to inclusion or even before in their available medical records. HZ cases were identified in HES using the International Classification of Diseases 10th revision (ICD-10) codes that appeared in the diagnosis fields. If HZ diagnosis codes were recorded in both HES and the CPRD, the earliest event date was considered as the onset date.

METHODS, Row 152-153: Please clarify how the patients were selected. My understanding is that HZ and IC patients were matched to HZ and IC-free patients, so the sample selection flow should be 1) selecting patients with HZ, 2) identifying IC and IC-free cohorts, and 3) matching IC-free cohort to IC cohort. But the paper reads like 2)□3) □1).

The reviewer is correct, we have modified the paragraph to read "The matched IC and IC-free cohorts were followed up from the index date until the earliest of the following events: transfer out of the

practice date, the last GP practice collections date, death date or the end of the study [Yanni et al. 2018].” Healthcare resource data associated with an incident HZ episode during the study follow-up were extracted for IC and Matched IC-free HES-linked individuals.”

DISCUSSION Row 248-251: The paragraph mentioned “the calculation of IC condition prevalence rates, HZ incidence rates”. But these outcomes were not assessed at all in the paper.

Please clarify.

The reviewer correctly points out that “the calculation of IC condition prevalence rates, HZ incidence rates were not presented in this paper. For clarification we’ve added “see Yanni et. al. for further detail on epidemiological outcomes [Yanni et al. 2018].”

DISCUSSION Row 283-285: This paragraph mentioned “higher incidence rates of HZ only”. Please provide a reference for this.

Generally, less than 30% of HZ patients develop PHN as such more than 70% have HZ only. Therefore, the incidence of HZ only is higher than the incidence of HZ with PHN.

DISCUSSION Row 287-298: This paragraph only cited the previous published study. The authors could elaborate how their study different from the previous studies and their contribution to the literature. Since the methods used in this study is quite straight forward, compared to some previous studies, the authors could use discuss the strength (large sample size?) of their paper here.

We’ve added the following “Insurance databases include not only the healthcare resource utilization but also costs. In the CPRD and HES Databases only the resource utilization is captured. As such the overall costs need to be calculated by assigning unit costs to the resource utilization. There are advantages of using the CPRD and HES in that the databases offer more diversity than might be observed using insurance databases, the latter of which may be somewhat limited by bias associated with factors such as age, race, and income.”

CONCLUSION Row 308-309. The conclusion stated that “In conclusion, individuals with IC conditions incurred higher healthcare utilization and costs than IC-free individuals.” Please put

more specifications on this conclusion, for example, in UK, patients seeking healthcare in general practices.

We’ve added “In conclusion, individuals with IC conditions seeking healthcare in the UK, incurred higher healthcare utilization and costs than IC-free individuals.”

References:

1. Yawn BP, Itzler RF, Wollan PC, et al. Health care utilization and cost burden of herpes zoster in a community population. Mayo Clin Proc 2009;84(9):787-94.

2. Li A, Chen SY, Burstin SJ, et al. Cost of Herpes Zoster in Patients With Selected Immune-Compromised Conditions in the United States. Open Forum Infect Dis 2016;3(2):ofw067.

3. Yanni EA, Ferreira G, Guennec M, El Hahi Y, El Ghachi A, Haguinet F, Espie E, Bianco V. Burden of herpes zoster in 16 selected immunocompromised populations in England: a cohort study in the Clinical Practice Research Datalink 2000-2012. BMJ Open. 2018 Jun 7;8(6):e020528.

4. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827-36. doi: 10.1093/ije/dyv098. Epub 2015 Jun 6.

VERSION 2 – REVIEW

REVIEWER	Qian Li US
REVIEW RETURNED	04-Dec-2018

GENERAL COMMENTS	Only some minor comments: Row 154, "Matched IC-free" don't need to be capitalized. Optional, I think the national representativeness of CPRD is one strength of the paper. Authors may want to mention it. Optional, some coding details are added (e.g. row 180 prodcode-packtype-quantity, row 169 CPRD tbl:therapy). Are these necessary? Suggestion excluding Supplementary Table 6 so the paper will focus on HZ-related costs. Authors response text already addressed my comment very well.

REVIEWER	Rafael Mikolajczyk Martin-Luther-University Halle-Wittenberg, Germany
REVIEW RETURNED	08-Mar-2019

GENERAL COMMENTS	Well executed, well described analysis. I have just a single comment on Limitation section - p. 17, line 330: I could not Access the paper by Schröder et al. but I doubt very much that the German Health insurance data have more details to classify IC into low and high - particularly they do not include laboratory parameters - but include medication. I would think that Information in CPRD is even better. So please recheck the references and state more precisely where is the difference.

REVIEWER	Francesco D'Amico PSSRU - Health Policy London School of Economics United Kingdom
REVIEW RETURNED	19-May-2019

GENERAL COMMENTS	Dear authors, I have enjoyed reading this paper that you submitted to the BMJ Open. I have been asked to review this article with a particular emphasis on the methods and analyses used, as the medical aspects are beyond my knowledge. I would like therefore to raise a few points that perhaps could be useful to address: - This paper estimates immuno-compromized (IC) condition prevalence rates, HZ incidence rates and HZ-related healthcare utilization and associated costs. It does provide information about different pattern of these quantities between IC and IC-free groups, but it is not explicit about whether these differences are significant (some of the tables present SDs but no mention is made about p-values or confidence intervals). I think this aspect could be discussed more extensively and perhaps added to the conclusions. - One objectives of the paper, in the way it is described in the conclusive section of the abstract, seems a bit tautological, as it states that people that are within the IC group are both more likely to suffer by HZ and to incur in higher HZ-related costs. Perhaps this latter concept could be rephrased by giving more emphasis to the extent/percentage by which costs are increased within the IC group. - When performing an observational study, it is important to acknowledge that the two groups being analysed are very likely to be heterogeneous, and that some form of matching or regression should be performed. This paper seems to have performed a form of matching, as it refers to "HES-linked Matched IC or IC-free cohorts". However, I could not find a lot of details on how this matching has been performed, and being clearler about this key-aspect would be improve the clarity of the paper. Some useful references on the subject could be the following ones: - Ruta Brazauskas and Brent R. Logan, "Observational Studies: Matching or Regression?", Biol Blood Marrow Transplant. 2016 Mar; 22(3): 557–563. Published online 2015 Dec 19. doi: 10.1016/j.bbmt.2015.12.005 - Herbert L. Smith, "Matching With Multiple Controls to Estimate Treatment Effects in Observational Studies", Sociological Methodology, Volume27, Issue1, 1997, Pages 325-353, https://doi.org/10.1111/1467-9531.271030. - The conclusions section should be expanded and made more adherent to the research question. Currently there are only two sentences, but they do not explicity mention the link between IC and HZ. I agree that this link has been mentioned throughout the whole paper, but in my opinion this section seems a bit incomplete.

REVIEWER	Iacopo Baussano International Agency for Research on Cancer, Lyon, France
REVIEW RETURNED	20-May-2019

GENERAL COMMENTS	This is an invited methodological Review of the manuscript bmjopen-2018-023502.R1entitled "HERPES ZOSTER RELATED HEALTHCARE BURDEN AND COSTS IN IMMUNOCOMPROMISED (IC) AND IC-FREE POPULATIONS IN ENGLAND: AN OBSERVATIONAL RETROSPECTIVE DATABASE ANALYSIS". The stated study aim is to estimate the healthcare resource utilization of HZ in selected immunocompromised populations and in an immunocompetent population aged ≥18 years of age in England and overall the manuscript addresses clearly this issue. I have however a couple of major comments. A) No statistical inference is attempted. Tables and figures report only descriptive figures (i.e. mean, medians, and standard deviations), nevertheless no attempt to assess the significance of observed differences is attempted. The authors should clarify the reasons of such a decision or provide estimates of uncertainty of their comparisons between immunocompromised and immunocompetent groups. B) Description of the sources of data (i.e., 1. CPRD GOLD 2014Q3; 2. HES Inpatient 2013Q3; 3. HES Outpatient data) The authors should provide in the main text a detailed summary of relevant characteristics of data sources and should clarify how these have been used to identify the immunocompromised and immunocompetent cohorts. This is important to design sensitivity analyses necessary to overcome the crucial limitations that "diagnoses were derived from administrative codes, which are recognised to be subject to miscoding or under-coding and are not validated against medical charts". This limitation is stated, but no attempt is made to assess its possible impact on the estimates and to limit its effect on such estimates. Lack of detailed information about the sources of data makes it also difficult to interpret key statements, such as "the population cohort linked to the Hospital Episode Statistics (HES) database was matched to a cohort of HES linked IC-free population (N=621,588), by age, gender and practice location". Why the source 1 (i.e. CPRD GOLD 2014Q3) could not be used to identify immunocompetent individuals? Why is it necessary to use HES? More details are needed for readers not familiar with these sources of data. C) The authors state "This study has several limitations", but only state (but do not assess) point B. Please clarify what are the other limitations and how are they addressed? Minor comments I am not sure why we need "IC-free" instead of "immunocompetent". Neologisms are meant to fill a semantic gaps, I am not sure what is expressed with "IC-free" that is not with "immunocompetent".

VERSION 2 – AUTHOR RESPONSE

Reviewers' Comments to Author:

Reviewer: 3

Reviewer Name: Qian Li

Row 154, "Matched IC-free" don't need to be capitalized.

Thank you for noticing this inconsistency

Optional, I think the national representativeness of CPRD is one strength of the paper. Authors may want to mention it.

Thank you for your suggestion, we have added text to the discussion as follows: “A strength of the CPRD is that it is considered to be broadly representative of the characteristics of patients and GP practices in the UK”. We have also added the reference Herrett et al. 2015 to support this statement.

Optional, some coding details are added (e.g. row 180 prodcode-packtype-quantity, row 169 CPRD tbl:therapy). Are these necessary?

We have deleted the text as suggested.

Suggestion excluding Supplementary Table 6 so the paper will focus on HZ-related costs. Authors response text already addressed my comment very well.

Thank you for your suggestion. We believe that readers may also pose this question, i.e. how does the non-HZ related healthcare resource utilization differ between IC and IC free subjects. As such we are happy to keep supplementary Table 6 to address potential questions.

Reviewer: 4

Reviewer Name: Rafael Mikolajczyk

Well executed, well described analysis.

Thank you for your kind words.

I have just a single comment on Limitation section - p. 17, line 330: I could not Access the paper by Schröder et al. but I doubt very much that the German Health insurance data have more details to

classify IC into low and high - particularly they do not include laboratory parameters - but include medication. I would think that Information in CPRD is even better. So please recheck the references and state more precisely where is the difference.

The reviewer makes a very valid point. For example, in the manuscript by Schroder et al. HIV subjects were classified as “High IC” from diagnosis until end of study period. Given that, for HIV, not all subjects will be immunosuppressed at any given time (i.e. CD4 counts <200) the classification of all HIV subjects as “High IC” is difficult to support. As such, we didn’t feel comfortable classifying individuals into high and low IC categories as we believed that “insufficient details are recorded in the CPRD and HES databases to allow adequate definition of patients’ severity of immunosuppression e.g. laboratory parameters, immunosuppressive medication details such as chemotherapy.”

Reviewer: 5

Reviewer Name: Francesco D'Amico

Dear authors, I have enjoyed reading this paper that you submitted to the BMJ Open.

Thank you for your kind words.

This paper estimates immuno-compromized (IC) condition prevalence rates, HZ incidence rates and HZ-related healthcare utilization and associated costs. It does provide information about different pattern of these quantities between IC and IC-free groups, but it is not explicit about whether these differences are significant (some of the tables present SDs but no mention is made about p-values or confidence intervals). I think this aspect could be discussed more extensively and perhaps added to the conclusions.

We agree with the reviewer that we performed these analyses as descriptive analysis only. Note in the protocol approved by Independent Scientific Advisory Committee (ISAC), we detailed that the analysis would be descriptive in nature only. We’ve completed the analysis as specified in the protocol. We were interested in describing the HZ-related costs in the 2 populations (i.e. IC and IC-Free). Our aim was not to demonstrate that there are significant differences between them. Although, given the large sample sizes included in this study it is highly likely that most of the comparison between IC and IC-free populations would be statistically significant. We edited the text in the conclusion to state that “In this descriptive analysis, involving a large representative national data source, the results suggest that individuals with IC conditions, were associated with higher HZ related healthcare utilization and costs than IC-free individuals”. Following the reviewer’s comments, we have also decided to add Standard Errors (SE) to tables 2, and 3 (i.e. replacing the standard deviations). This allows the reader to generate informal 95% confidence intervals, i.e. using $\mu \pm 1.96 * SE$. Note using these 95% confidence intervals may not be entirely appropriate as the variables may not be normally distributed. The SE (and CIs) may also be used by the reader as a guide to inform statistical significance. See also response to reviewer 6 below.

One objectives of the paper, in the way it is described in the conclusive section of the abstract, seems a bit tautological, as it states that people that are within the IC group are both more likely to suffer by HZ and to incur in higher HZ-related costs. Perhaps this latter concept could be rephrased by giving more emphasis to the extent/percentage by which costs are increased within the IC group.

We have edited the text to state “Individuals with IC conditions, have a high burden of HZ, associated with an increased risk of HZ and high HZ-related healthcare costs.” We believe that this is an important point as some authors define burden as incidence only, but it is also important to take into account costs (or severity) also.

When performing an observational study, it is important to acknowledge that the two groups being analysed are very likely to be heterogeneous, and that some form of matching or regression should be performed. This paper seems to have performed a form of matching, as it refers to "HES-linked Matched IC or IC-free cohorts". However, I could not find a lot of details on how this matching has been performed, and being clearler about this key-aspect would be improve the clarity of the paper.

Indeed, in this study we performed a matching. We’ve edited the text to state: “The CPRD IC population cohort was linked to the Hospital Episode Statistics (HES) database and matched, using a 1:1 ratio, to a cohort of CPRD-HES linked IC-free population (N=621,588), by age, gender and practice location” As suggested we’ve added the reference: “Ruta Brazauskas and Brent R. Logan, "Observational Studies: Matching or Regression?", Biol Blood Marrow Transplant. 2016 Mar; 22(3): 557–563.” The supplemental text provides more detail on the Matching.

- The conclusions section should be expanded and made more adherent to the research question. Currently there are only two sentences, but they do not explicity mention the link between IC and HZ. I agree that this link has been mentioned throughout the whole paper, but in my opinion this section seems a bit incomplete.

Thank you for spotting this omission, i.e. we did not explicity mention the link between IC and HZ. We have edited the section as follows: “Immunosuppression is known to be associated with an increased risk of HZ in the UK [Forbes et al. 2014. Yanni et al. 2018]. In this descriptive analysis, involving a large representative national data source, the results suggest that individuals with IC conditions, were associated with higher HZ related healthcare utilization and costs than IC-free individuals”. Note in a previous version of the manuscript we attempted to put the results into context given the available vaccines, however given our affiliation we understand that this may have appeared as promotional.

Reviewer: 6

Reviewer Name: Iacopo Baussano

The stated study aim is to estimate the healthcare resource utilization of HZ in selected immunocompromised populations and in an immunocompetent population aged ≥18 years of age in England and overall the manuscript addresses clearly this issue.

I have however a couple of major comments.

A) No statistical inference is attempted.

Tables and figures report only descriptive figures (i.e. mean, medians, and standard deviations), nevertheless no attempt to assess the significance of observed differences is attempted. The authors

should clarify the reasons of such a decision or provide estimates of uncertainty of their comparisons between immunocompromised and immunocompetent groups.

See also response to reviewer 5 above. From the beginning of the study we planned to analyse the data descriptively only. When we started developing the protocol we acknowledged that there was a lack of data regarding immunocompromised populations and even a lack of a definition of IC populations. We selected 16 IC conditions in our definition of an IC population but perhaps other researchers would select different IC conditions. As such our study is exploratory in nature and was not intended to be definitive. Nevertheless, following the reviewer's comments, we have also decided to add Standard Errors (SE) as estimates of uncertainty to tables 2, and 3 (i.e. replacing the standard deviations). This allows the reader to generate informal 95% confidence intervals (Cis), i.e. using $\mu \pm 1.96 * SE$. Note using these 95% CIs may not be entirely appropriate as the variables may not be normally distributed. The SE (and CIs) may also be used by the reader as a guide to inform statistical significance.

B) Description of the sources of data (i.e., 1. CPRD GOLD 2014Q3; 2. HES Inpatient 2013Q3; 3. HES Outpatient data) The authors should provide in the main text a detailed summary of relevant characteristics of data sources and should clarify how these have been used to identify the immunocompromised and immunocompetent cohorts. This is important to design sensitivity analyses necessary to overcome the crucial limitations that "diagnoses were derived from administrative codes, which are recognised to be subject to miscoding or under-coding and are not validated against medical charts". This limitation is stated, but no attempt is made to assess its possible impact on the estimates and to limit its effect on such estimates.

We have provided more detail in the supplemental text on the datasets used, i.e. the characteristics of the databases used and how these have been used to identify the immunocompromised and immunocompetent cohorts. We highlighted "diagnoses were derived from administrative codes "as a limitation however this is a limitation of many database studies. For example, a study in The Netherlands reported that using "free-text" fields in addition to administrative codes they identified additional incident cases of varicella and herpes zoster [Pierik et al. 2012]. We have added this latter reference to the manuscript. In the CPRD database the GP is also able to make additional uncoded notes and observations about patients as free text. This often contains identifiable information and is not part of the standard database available to researchers. Although we recognize that this is a potential limitation we believe that the diagnostic codes reported in the CPRD are quite accurate as have been documented elsewhere [Herrett et al. 2015]. This is reflected in the HZ incidence rates reported for this study by Yanni et al. 2018 where the values are very much in line with a systematic review performed by Kawai et al. and the UK Green Book, Chapter 28a which reports incidence values of HZ in the UK general population. As such we consider this as not a minor limitation of our study.

Lack of detailed information about the sources of data makes it also difficult to interpret key statements, such as "the population cohort linked to the Hospital Episode Statistics (HES) database was matched to a cohort of HES linked IC-free population (N=621,588), by age, gender and practice location". Why the source 1 (i.e. CPRD GOLD 2014Q3) could not be used to identify immunocompetent individuals?

We have now provided more detail on the sources of data in the supplementary text. We've edited the text to clarify: "The CPRD IC population cohort was linked to the Hospital Episode Statistics (HES) database and matched, using a 1:1 ratio, to a cohort of CPRD-HES linked IC-free population (N=621,588), by age, gender and practice location" As such the source 1 (i.e. CPRD GOLD 2014Q3)

is used to identify immunocompetent individuals. We added a sentence at the end of the methods section as follows “Further details, including information on the IC populations included, ICD-10 codes for HZ and PHN, and unit healthcare costs are provided in the Supplementary Text.” We also edited the results section to state “The CPRD-HES-linked matched IC and IC-free population cohorts (n=621,588 each)”

Why is it necessary to use HES?

Although most HZ cases would be diagnosed in the GP practice and recorded in the CPRD, HES-linked cohorts were selected for analysis to identify HZ patients who might be admitted to hospitals because of HZ complications. Hence, HES-linked data represented the most complete dataset available. In addition, when calculating healthcare related costs it is important to include all the relevant costs, i.e. in particular the hospitalization costs.

More details are needed for readers not familiar with these sources of data.

C) The authors state "This study has several limitations”, but only state (but do not assess) point B. Please clarify what are the other limitations and how are they addressed?

We have added another limitation as follows: “In this study we selected 16 IC conditions in our definition of an IC population but perhaps other researchers would select different IC conditions. As such our study is exploratory in nature and was not intended to be definitive.”

Minor comments

I am not sure why we need "IC-free" instead of “immunocompetent”.

Neologisms are meant to fill a semantic gaps, I am not sure what is expressed with "IC-free” that is not with “immunocompetent”.

When developing the protocol, the statistical analysis plan and statistical tables we used the terminology “IC” and “IC-Free”. For consistency, we have continued to use the terminology also in our previous manuscript: Yanni et al.. BMJ Open 2018;8:e020528, and in this manuscript.

References:

1. Yawn BP, Itzler RF, Wollan PC, et al. Health care utilization and cost burden of herpes zoster in a community population. Mayo Clin Proc 2009;84(9):787-94.

2. Li A, Chen SY, Burstin SJ, et al. Cost of Herpes Zoster in Patients With Selected Immune-Compromised Conditions in the United States. Open Forum Infect Dis 2016;3(2):ofw067.

3. Yanni EA, Ferreira G, Guennec M, El Hahi Y, El Ghachi A, Haguinet F, Espie E, Bianco V. Burden of herpes zoster in 16 selected immunocompromised populations in England: a cohort study in the Clinical Practice Research Datalink 2000-2012. BMJ Open. 2018 Jun 7;8(6):e020528.

4. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827-36. doi: 10.1093/ije/dyv098. Epub 2015 Jun 6.

5. Pierik JG, Gumbs PD, Fortanier SA, Van Steenwijk PC, Postma MJ. Epidemiological characteristics and societal burden of varicella zoster virus in the Netherlands. BMC Infect Dis. 2012 May 10;12:110. doi: 10.1186/1471-2334-12-110.

6. Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. BMJ Open 2014;4:e004833.

7. Shingles (herpes zoster): the green book, chapter 28a https://www.gov.uk/government/publications/shingles-herpes-zoster-the-green-book-chapter-28a

VERSION 3 - REVIEW

REVIEWER	Francesco D'Amico PSSRU, London School of Economics, United Kingdom
REVIEW RETURNED	25-Jun-2019

GENERAL COMMENTS	Dear authors, thank you for addressing the comments included in my previous review. I am satisfied with the changes.