Author's response to reviews

Title: Herpes Zoster surveillance using electronic databases in the Valencian Community, Spain.

Authors:

Nuria NMT Morant-Talamante (morant_nur@gva.es)
Javier JDD Diez-Domingo (diez_jav@gva.es)
Sergio SMU Martínez_Ubeda (martinez_serube@gva.es)
Joan JPB Puig-Barberá (puig_joa@gva.es)
Sara SAS Alemán-Sánchez (aleman_sar@gva.es)
Lina LPB Pérez-Breva (perez_lin@gva.es)

Version: 2 Date: 9 August 2013

Author's response to reviews: see over
Dear editors,

We really appreciate the efforts and suggestions of the reviewers, which undoubtedly have contributed to improve our work. We have incorporated most of the recommended insights as described below.

**REVIEWER'S REPORT 1**

**REVIEWER: ANGEL GIL**

*Should improve the quality of the tables because there is too much information and not appreciated clearly. Maybe they should point out or mark the most relevant aspects*

Some changes have been made in the tables. However, we give this amount of information so that researchers with interest in the field can compare their results. Similar tables are published in other papers.

*You could make a brief analysis of the disease costs to analyze the potential impact of vaccination.*

Although this is far from the objectives of the study, we have included some additional data describing the impact of the disease.

*Enlarge the literature with disease burden studies done in and out of Spain.*

Due to the large number of studies assessing the epidemiology of HZ throughout the world, we mainly included studies that used electronic databases. We have added all papers found regarding the burden of disease in Spain.

**REVIEWER'S REPORT 2**

**REVIEWER: KAITLIN RAINWATER LOVETT**

**MAJOR COMPULSORY REVISIONS**

**BACKGROUND**

*In the 2nd sentence of the 2nd paragraph you discuss how HZ incidence could change due to a future demographic transition in Europe, but you provide no references or explanation of how HZ incidence related to the age of the population. This information is critical to this argument as well as to your findings showing a strong age component in HZ epidemiology.*
Explain and provide references for prior epidemiological evidence of an age effect on HZ incidence!

Changes were made in order to better explain the basis of the study. References indicating an increased incidence with older age are now included.

The 3rd paragraph needs to be expanded and re-phrased as this is the motivation for this study! Perhaps be more specific about the results of the mathematical models. Explain the biological mechanism and plausibility regarding decreased cellular immunity following varicella vaccination. There are parallels with other diseases you can use here. Additionally, is there any epidemiological evidence demonstrating an increase in disease due to a lack of specific types of immunity after vaccination?

Changes were made for better understanding of this paragraph. We stated that the aim of our study is to assess the incidence of the disease before HZ vaccination with a system that allows us to estimate the impact of HZ vaccination once it is included in the Spanish NIP. The results of this mathematical model are mentioned as a possibility that should be carefully monitored, but we do not expect any impact of varicella vaccine on the incidence of HZ in Valencia. This is now explained in the following paragraph:

“...In the Valencian Community (Spain), varicella vaccination is recommended and funded at 12 years of age in those subjects who have not been in clear contact with the virus. This helps avoid severe cases in adults with no impact upon circulation of the virus. About 30% of toddlers also receive the vaccine after their pediatrician’s recommendation (new reference) With this low coverage figure, the virus is circulating within the population and no effect upon HZ is expected.”

We tried to stress that the study interest relies on the future HZ vaccine, and not on the present varicella vaccine program.

The end of the 4th paragraph references “unpublished public health data”. Typically, unpublished data is cited along with the name of the author/s who are developing this research, but “public health data” implies you received this information from a public health agency or other organization. Where did you get this from? Please describe in the Methods section or cite as a “personal communication” alongside the person’s name who provided the data.

Varicella vaccination coverage was provided by the Department of Health of the Generalitat Valenciana (Valencian Community Government), through the Directorate for Public Health. Data are unpublished.
Since changes were made in this paragraph, and we do not want to emphasize that chickenpox vaccination could alter HZ incidence in adults, we added another recent reference.

METHODS

The description of the databases that were combined for analysis should be written in a consistent format. Currently, SIA and GAIA are basically in an outline format, while the others are described in a paragraph format. An outline also is used for the identification of incident cases of HZ. Perhaps you can combine these two sections into paragraph format where you describe a database and then describe what type of information you extracted from this specific database for your study.

Changes have been made in order to better explain the Abucasis database.

In “For identification of incident HZ cases” section:
- In SIA, please elaborate on what is incorporated into ICD-9-CM 053 codes. The reader shouldn’t have to assume it is HZ or look this up if they don’t know. In CMBD, why use only the first to ninth listed diagnoses?
- In GAIA, these drugs (acyclovir, famciclovir and valacyclovir) can be prescribed for several types of herpes infections including genital and oral herpes. If you used this prescription information and did not require a diagnosis, you are including individuals in your surveillance that likely does not have HZ.

- A short explanation of ICD-9-CM 053 code related diagnoses has been given. We consider it too long to give a description of almost 20 subcodes, all related to HZ. Also, in Table 4, the description of the most frequent HZ CD-9-CM is included.
- The CMBD database at that time only provided 9 listed diagnoses. At present over 30 diagnostic codes are incorporated in CMBD, which is why we stated that we included all 9 diagnoses.
- GAIA: we only took into account antivirals prescribed at high doses which in Spain are only licensed as HZ treatment. However, as described in the Results section, the reliability of this parameter was so low that we discarded any case with antiviral treatment and no HZ codification.

Where did you obtain information to support the correction factors used here? Please provide a reference as well as a more detailed description of how these correction factors were used in the weighting of HZ incidence rates.

As mentioned in the paper, Abucasis (SIA-GAIA) was progressively incorporated in all the regional health system. Abucasis managers provided us the number (and name) of the Health Centers with the system implemented per month and therefore the proportion of the population of Valencia included each year. Thanise are unpublished data, and we expect that this study will serve as the reference for future epidemiological studies using the databases. We have to take into account that the Abucasis system was not established for
epidemiological studies, and we had to obtain some information through personal interview.

**How was the subset of 550 medical records chosen for review by a physician?** You state it was random but what strategy/algorithm did you use to ensure “randomness”? The method of selection can bias your assessment of the validity and reliability of the databases.

Random method of the MySQL system was used as now described in the paper:
"The random selection for quality control of coding was performed using its “rand” function”.

In the first paragraph of Methods, you use the dates 1st January 2007 to 20th December 2010 for identifying HZ cases but in the medical records review paragraph you use 1 June 2006 to 31 March 2011. Is this a typographical error, or did you review medical records of a larger calendar range for a reason, perhaps to determine recurrence? If so, you should state this.

We used a longer period of time to determinate recurrence, so we added this information to state this.
"In order to determine recurrence, each medical contact in the Abucasis database from 1 June 2006 to 31 March 2011 was evaluated.”

**RESULTS**

The PPV for antiviral treatment was extremely low, which likely reflects the inclusion of non-HZ herpes infections. Is that why this is included here? The reasoning for this inclusion should be more explicit.

The following sentence has been added:
"The PPV for HZ diagnosed only by high-dose antiviral prescription was 26% (95%CI: 17.7-35.7), therefore, no case identified only by the prescription of antivirals was considered as an HZ case.”

You identified a total of 85,586 cases of HZ but 3,330 persons met the criteria for recurrence, so why do you continue to use 85,586 persons throughout the Results section? Their recurrent diagnoses are, by definition, not incident.

We have made changes for better understanding. We identified 85,586 first (incident) cases, and of these, 3330 persons had more than one episode.

“Our criteria of recurrence were met by 3300 of them. After chart review, the positive predictive value (PPV) for HZ case definition was 92.7% (95%CI 89.1-95.4), versus 55.1% for recurrent HZ cases (95%CI 47.0-63.0). Due to the low
PPV of the recurrence filter, only the incident HZ cases were included in our estimations.

**Under “HZ incidence”, you claim there are “varying [incidence] figures in different years but without a clear pattern” and reference Figure 1; however, Figure 1 shows an incredibly clear pattern of increasing incidence with increasing age, regardless of the year. The differences in year are minimal – if you claim they are different, do you have a statistical test to support this? If not, I would not interpret them as such.**

The paragraph has been rewritten. It was a language problem. We meant that small changes in the epidemiology over the years occurred but there was no increasing or decreasing pattern.

“Over the four-year study period, we identified 85,586 persons with incident cases of HZ requiring medical care, which correspond to an incidence in all age groups of 4.60/1000 person year (PY) (95%CI: 4.57-4.63). The figures remained rather constant over the years: 4.58/1000 PY (95%CI: 4.51-4.65) in 2007, 4.89 (95%CI: 4.83-4.95) in 2008, 4.67 (95%CI: 4.61-4.73) in 2009, and 4.29 (95%CI: 4.24-4.35) in 2010 (Figure 1).”

**Differences in HZ incidence between men and women are clearly described in multiple locations of the Results section but the data are never stratified by sex. This is an important stratification if they truly differ. By how much do the incidence rates differ? Is this a statistically significant difference? At the least, I suggest stratifying Table 2 to show these differences.**

When there are sex differences, these are shown by the different confidence intervals. Tables in general have much information, and we wanted to mention some age differences but not emphasize them. We have included the overall incidence by sex and age group in Table 1. Table 3 includes these differences in hospitalized subjects.

**DISCUSSION**

If sex and age differences are strongly associated with HZ incidence rates (as you describe in the Results), why do you present a pooled incidence rate of 4.60 per 1000 PY? This comment is also in reference to the #1 years old bars in Figure 2. There are definitely differences by age group so why show a summary measure?
4.60 per 1000 PY is the overall incidence in our population. It is an important number that can be used for comparison with other studies. We have now added some comments on the higher incidence in women.

*If medical personnel were not trained in ICD-9-CM coding, how can you trust that this data is reliable? This is a huge limitation and needs to be thoroughly discussed.*

We have explained things better in the Methods and Discussion sections. In the Discussion we wrote:

“One of the limitations of the study is that the Abucasis database was not developed for medical billing purposes, and this may imply the obtainment of less detailed information. Although ICD-9-CM codes are routinely used, general practitioners (GP) may be unaware of its importance. To overcome this, codification is performed in simple way – the system presenting a list of possible ICD-9-CM codes after a plain text is written. In the case of Zoster, which is a common diagnosis, with no synonyms, ICD-9 coding is relatively easy…”

*You state that the high burden of care in primary care clinics may bias the results. I agree with this but please offer an explanation as for why you think this might produce a bias.*

The following paragraph has been edited:

“. In order to assess the reliability of HZ codification, a random review of medical records showed good matching between ICD-9-CM coding and a real episode of HZ. A total of 7.7% of the reviewed notes with the diagnosis of Herpes Zoster did not meet the full requirements for being considered a case. Most of these situations were due to a lack of description of the lesions. We considered that most of these cases could be HZ, but the GP did not provide sufficient written details, possibly because of the great care burden found in primary care, which prevents entering detailed information in medical notes. This consequently may have understated HZ case confirmation.”

We now consider how busy clinics may affect the study.

*The reliability of the data is not confirmed by comparing Valencia to the rest of Spain. Moreover, if you want to claim this is a reliable system, you need to show the data to the reader.*

We have added some data from Spain in the Discussion.

*The overall increase in HZ with increasing age perhaps reflects immunosenesence, which you do not mention. The peak in incidence among individuals >95 years old may also be due to small sample size in this age range, which would create incidence estimates with very large confidence intervals. Less likely but also a possibility is a survivor effect - individuals who developed incident HZ later in life simply lived longer than those who did not develop HZ. More of the Discussion should be dedicated to evidence supporting a link between low varicella vaccination*
rates and an increase in HZ incidence, which is the motivation for the paper.

Additional discussion has been added.

MINOR ESSENTIAL REVISIONS

BACKGROUND

The 3rd sentence of the 1st paragraph requires a reference ("Approximately 14% of patients...")

Clarified.

The last sentence of the 1st paragraph needs a space between the last word and the reference.

We placed the space.

In the 2nd paragraph, the phrase “1000 persons-year (PY) (annual events per 1000 population)” is awkward. This is most often expressed as, “1000 person-years (PY)". Alternatively, if the authors are concerned about interpretation by the reader, it may be more concise and clear to phrase as, “1000 persons per year (PY)".

We omitted the clarification of persons-year, since it is a common expression in epidemiological studies and might not need further explanations.

The 3rd sentence of the 4th paragraph is unclear. What is the “official routine paediatric immunization calendar for children aged >=11 years”? Does this mean children who are older than 10 years should be vaccinated against varicella if they have not previously received this vaccine?

Yes, that is the meaning of this sentence, but we changed the expression for better understanding. It was also included in the Introduction.

With respect to the efficacy of the Zostavax – be more specific here. What was the efficacy? Can you provide a percentage of individuals who responded appropriately to the vaccine?

We included results of different studies on the efficacy of Zostavax®.

The last sentence/paragraph should have the word “and” placed between “...to December 2010)” and “to estimate the reliability...”.

The word “and” has been placed.
METHODS

The last sentence of the “Setting and study population” section is unclear. Is this intended to introduce medical records review to the reader, which is explained in more detail later? If so, you can delete this sentence and introduce when you present more detail.

This sentence was initially placed in this section in order to explain that after reviewing medical records, more filters were applied. Since we made changes in the whole section, we deleted from here.

The heading Abucasis electronical healthcare database – electronical is not a word in English. I think you mean “electronic”.

Correction has been changed.

What are “clinical notes”? Are these the notes written by physicians in an individual’s chart? Perhaps you can re-phrase this as “relevant clinical information from physician’s notes”? It is unclear what you specifically gained from these and how the information you obtained was systematically used - this needs more description.

We tried to better the Abucasis database describe (see above).

Standard English writing style typically specifies that you write a phrase and put the acronym in parentheses afterwards. There are numerous occasions where you write the acronym and put the phrase in parentheses. This is difficult for the reader. Additionally, you do not need to write the phrase of an acronym in both English and Spanish since it is clear this study occurred in an area where English is not the primary language.

Appropriate changes have been made.

In the 2nd paragraph of the Abucasis electronical healthcare database section, you should use semi-colons, such as: “…any ambulatory medical annotation; laboratory, imaging or additional tests; physician-liaisons, and medical reports.”

Semicolons have been placed.

In the 2nd sentence of the 4th paragraph of the Abucasis electronical healthcare database section, the word “also” should be placed after the word “databases” so it reads: “The databases also can be linked…”

These changes have been made.

In the GAIA paragraph in the Identification of incident HZ cases section, the word “nor” should be replaced with “or”.


Done.

*The initials of the author who reviewed the medical records should be placed after “reviewed by a physician”.*

Done.

You provide a reference but not a description of the cohort that was used for comparison. Please include a short one or two sentence description to give the reader context for this comparison.

The study was partially mentioned with reference in background section, but we gave a short description in this section.

**Furthermore, HOW did you compare the cohort to the surveillance results?**

We have added the following sentence:
“Differences were assessed from the overlapping of confidence intervals“.

*The use of statistical software is described but what statistical analyses did you perform? Was this only used to generate 95% CIs? More importantly, how were 95% CIs calculated?*

We have added the following paragraph: “The exact 95% confidence interval (CI) for the incidence rates was calculated on the basis of a normal distribution. We considered differences between males and females or among age groups when the confidence intervals did not overlap.”

**RESULTS**

*Overall, there are several inconsistencies in the writing and the format could be improved. Sometimes you used commas in large numbers and sometimes you did not (85586 vs 85,586). Typically, a paragraph consists of 3 or more sentences. There are several instances of one-sentence paragraphs.*

We corrected this.

*How did you calculate PPV? Why discuss only PPV and not sensitivity and specificity?*

Only clinical notes with the ICD-9 codes of HZ were reviewed; therefore, we had no possibility of calculating sensitivity and specificity. However, we could
calculate the PPV of the codes after assessing whether the HZ codified notes fulfilled our criteria for HZ.

**A comparison between #15 year olds and #50 year olds is made. Does the #15 year old group really imply 15-49 year olds? If not, individuals who are #50 years are included in both groups.**

We made this comparison, since individuals ≥ 15 years of age are not in the pediatric age and moreover, this age group was needed to compare our results with those of the prospective study. We included the term “nonpediatric age” as explanation.

Subjects aged ≥ 50 years are the target population for an HZ vaccine, which is why we made this distinction. We have also explained this point. These age-cutoff points are needed in order to make comparisons with other European studies in discussion.

**Please be consistent in the number of significant digits that you include (26.2 vs 17.14, for example). Sometimes it’s 2, 3, 4 or more.**

We have reviewed and corrected this.

**The phrase: “Most codified HZ diagnosis: ICD-9-CM 053.9” is not a sentence and appears as its own paragraph. Change formatting and give context if you believe this is an important piece of information.**

This typing mistake has been corrected.

**Last paragraph, end of 2nd sentence: “...although the confidence intervals approached in both studies.” Approached what? I believe this is missing a number.**

We have changed the word to “overlapped”.

**DISCUSSION**

**The word “limit” is used when you mean “limitation”.**

Done.

**What is protein chain reaction? Do you mean polymerase chain reaction (PCR)?**

We have changed the sentence including a general term, since there are other laboratory tests to confirm HZ.

**Please provide the reference in the sentence that states, “...as the prospective study authors discussed...”.**

Done.
DISCRETIONARY REVISIONS

BACKGROUND

The 6th paragraph on the use of electronic medical records is not necessary. I would suggest putting the 2nd sentence in the “Abucasis electronical healthcare database” section of the Methods.

We have divided it between Methods and Discussion.

If you delete the 6th paragraph, the final paragraph/”purpose of the study” sentence can be combined with the previous paragraph. This sentence needs the word “and” inserted between “…December 2010), and to estimate…”, and could benefit from being split into two sentences. To be more clear, you can outline the specific purpose of the study in one sentence (epidemiology of HZ) and another sentence explaining the long-term goal of this research using the last phrase (“in order to create surveillance tools…”).

We have integrated the remaining sentence with the next paragraph.