Reviewer's report

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

Version: 1 Date: 24 April 2013

Reviewer: Kaitlin Rainwater Lovett

Reviewer's report:

Reviewer: Kaitlin Rainwater Lovett

Morant-Talamante et al. assessed the incidence and epidemiology of herpes zoster in Valencia, Spain, using electronic medical records' systems that recorded ICD-9-CM codes and were confirmed by physician chart review. The question posed by the authors was well-defined; however, the authors stated the motivation for understanding HZ epidemiology is related to potential increases in HZ incidence due to decreased cellular immunity after varicella vaccination relative to natural exposure. They provide no epidemiological evidence to support this claim, except mathematical models, and little explanation of the biological mechanism through which this would occur, but admit this is a controversial hypothesis. As the basis for the study, this requires more thorough explanation in the Background as well as in the Discussion about the implications of the study results on the epidemiology of HZ in the future. The Methods are not well-described, leaving the reliability of the data to be unclear. The writing could benefit from a thorough review of spelling, grammar and re-phrasing.

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!

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?

The end of the 4th paragraph references “unpublished public health data”. Typically, unpublished data is cited along with the name of the author/s who are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.

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.

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 do not have HZ.

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.

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

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.

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.

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.

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.

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.

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 #15 years old bars in Figure 2. There are definitely differences by age group so why show a summary measure?

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.

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 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.

The overall increase in HZ with increasing age perhaps reflects immunosenescence, 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.

Minor Essential Revisions

BACKGROUND

The 3rd sentence of the 1st paragraph requires a reference (“Approximately 14% of patients…”)

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

The 3rd sentence of the 4th paragraph is unclear. What is the “official routine pediatric 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?

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?

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

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.

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

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.

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.

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-liasons, and medical reports.”

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…”

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

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

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

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?

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.

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

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.

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.

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.

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

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

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

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

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.

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…”).

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.