Author's response to reviews

Title: Relationship between efficiency and clinical effectiveness indicators in an adjusted model of resource consumption: a cross-sectional study

Authors:

Concepcion Violan-Fors (cviolan@idiapigol.org)
Oleguer Plana-Ripoll (oplana@idiapigol.org)
Quintí Foguet-Boreu (qfoguet@idiapigol.org)
Bonaventura Bolíbar (bbolibar@idiapigol.org)
Alba Aguado (alba.aguado@sanitatintegral.org)
Ruth Navarro-Artieda (rnavarro.germanstrias@gencat.cat)
Soledad Velasco-Velasco (svelasco@bsa.cat)
Antoni Sicras-Mainar (asicras@bsa.cat)

Version: 2 Date: 29 November 2012

Author's response to reviews: see over
Reviewer's report
Title: Relationship between efficiency and clinical effectiveness indicators in an adjusted model of resource consumption: a cross-sectional study
Version: 2 Date: 30 October 2012
Reviewer: Lawrence Fulton

Reviewer's report:
The authors argue that ACGs are valid, because they have been used but then concur that definitions do not match those expected from health economics theory. If measures are not congruent with theory, then the authors have a conundrum. They are not measuring what is purported to be measured.

Our response to the reviewer's first-round comments was this: We agree with the reviewer that the efficiency indicators (referring to the use of resources) used by The Johns Hopkins ACG® System can lead to confusion with Efficiency as defined in Health Economics Theory. We are aware that economists make distinctions between technical, cost, and allocative efficiency. The extent to which the ACG® System contributes to achieving these efficiencies depends on how they are applied for payment purposes, which helps to determine the incentives for efficient behaviour in primary care.

Our point in making that statement is to make clear we are not attempting to apply classic Health Economics Theory to the questions addressed in this article. We are using a specific instrument, validated in the Spanish population, to assess the use of resources in a universal public healthcare system. This is the novel aspect of the research presented here: the addition of a measure of effectiveness to the existing ACGs (whose use was validated for Spain in a study published in BMC HSR in 2009 and cited in the current revision).

Even though we can concede that some confusion is possible for the reader, it is also possible to use a different lens to view this question of definitions -- and the publication of differing points of view is important to scientific progress. Our point of view is that the efficiency behaviour we want to measure is a comparison of the costs, number of visits, and episodes (i.e., a disease diagnosis that might require one or more visits) related to a particular type of ACG.

We could argue that this approach is a measurement of efficiency behaviours because it shows us how the same ACG can consume more or fewer resources (visits, costs or episodes). Measuring or comparing the capacity to do more with less -- in this case resources defined as visits, episodes and total costs -- is another way of measuring “efficiency” as defined in Health Economics Theory. Therefore, although we can concur that confusion may exist we cannot agree with the relevance of the comment that “definitions do not match those expected from health economics theory.”

It is incumbent on the authors to justify validity for their own study. The use of measures in previous studies does not make these measures valid. The
establishment of validity through testing needs to be detailed, not provided by a [2] or [3]. "They were shown to be valid" does not add to anybody's understanding.

**We have very clearly stated in this revision:** The ACG case-mix system has been validated in Spain by a 2005 retrospective, multi-centre study of 81,873 patients (30).

Succinctly summarize studies rather than use quotation numbers. I would suggest that most of these citations have nothing to do with validation but show use. Use does not equal validation.

Most of the studies sourced now in the conclusion session used the ACG measure without validation. The assumption was validity. Instead, they rely on the fact that ACGs are assumed to be valid because widely used ("the world is flat."). Instead of referring a reader to a Spanish and Swedish source, summarize the results of those studies, as validity is the authors' responsibility to support, not the readers.

**We reproduce here for the reviewer's convenience the abstract for the Swedish study – which, as its title indicates, was a validation study -- published in BMC Public Health:**

**Validating the Johns Hopkins ACG Case-Mix System of the elderly in Swedish primary health care**

**Background**
Individual based measures for comorbidity are of increasing importance for planning and funding health care services. No measurement for individual based healthcare costs exist in Sweden. The aim of this study was to validate the Johns Hopkins ACG Case-Mix System's predictive value of polypharmacy (regular use of 4 or more prescription medicines) used as a proxy for health care costs in an elderly population and to study if the prediction could be improved by adding variables from a population based study i.e. level of education, functional status indicators and health perception.

**Methods**
The Johns Hopkins ACG Case-Mix System was applied to primary health care diagnoses of 1402 participants (60–96 years) in a cross-sectional community based study in Karlskrona, Sweden (the Swedish National study on Ageing and Care) during a period of two years before they took part in the study. The
predictive value of the Johns Hopkins ACG Case-Mix System was modelled against the regular use of 4 or more prescription medicines, also using age, sex, level of education, instrumental activity of daily living- and measures of health perception as covariates.

**Results**

In an exploratory bi-plot analysis the Johns Hopkins ACG Case-Mix System, was shown to explain a large part of the variance for regular use of 4 or more prescription medicines. The sensitivity of the prediction was 31.9%, whereas the specificity was 88.5%, when the Johns Hopkins ACG Case-Mix System was adjusted for age. By adding covariates to the model the sensitivity was increased to 46.3%, with a specificity of 90.1%. This increased the number of correctly classified by 5.6% and the area under the curve by 11.1%.

**Conclusion**

The Johns Hopkins ACG Case-Mix System is an important factor in measuring comorbidity, however it does not reflect an individual's capability to function despite a disease burden, which has importance for prediction of comorbidity. In this study we have shown that information on such factors, which can be obtained from short questionnaires increases the probability to correctly predict an individual's use of resources, such as medications.

Similarly, the Spanish reference is a validation study, as the title indicates. We provide here the abstract in English, as the article was published in Spanish.


**ABSTRACT**

**Purpose:** To validate the Johns Hopkins ACG case-mix system used in various primary and specialized care centers attending a defined population in Spain.

**Methods:** A retrospective, multicenter study was carried out by applying the ACG case-mix system to the clinical records of patients attending five primary care teams and two hospitals over a 1-year period in 2005. The main measurements were dependent variables (visits, episodes, primary care costs, and total costs), and morbidity. The determination coefficient (R2; p<0.05) was used to measure the explained variability.

**Results:** A total of 81,873 patients were included with a mean (standard deviation) number of 4.8 (3.5) episodes and 8.0 (8.1) visits/patient/year. The
explained variance (R2) of ACG classification was 73.1% (75.5% log transformation) for episodes, 43.2% (54.0% log transformation) for visits, 19.6% (54.8% log transformation) for primary care costs, and 22.7% (48.3% log transformation) for total costs (p<0.001).

**Conclusion:** The ACG system classified a defined population on the basis of morbidity and individual resource consumption. Moreover, the ACG system was useful to assess the clinical (comorbidity) and economical information of each center.

**Key words:** Adjusted clinical groups (ACG). Resource use. Management. Information system.
Reviewer's report

Title: Relationship between efficiency and clinical effectiveness indicators in an adjusted model of resource consumption: a cross-sectional study
Version: 2 Date: 22 October 2012
Reviewer: Wui-Chiang Lee

Reviewer's report:
The authors have made significant revision in response to my, as well as another Reviewer’s, suggestions in this revised manuscript. In general, this manuscript is much better than the previous one, especially in the following areas:

1. The authors and the ACG Study Group in Spain have published a series of study results in terms of the validity and reliability of the ACG case-mix adjustment system based on their local database. I have no doubt of their capability and expertise in ACG use. However, the authors adopted many indices in this manuscript that not all reviewers are familiar with, such as EI, RI, and SI. I believe that the readers of the BMC Health Service Research will face the same barriers in their first time reading. The revised manuscript has elaborated more on the terminology in the Method section.

Thank you for your comments. With respect to the use of the EI, RI and SI indicators we would like to highlight the following points:

The efficiency index was previously used in a publication in BMC Public Health by one of the co-authors. [Sicras-Mainar A et al. The relationship between effectiveness and costs measured by a risk-adjusted case-mix system: multicentre study of Catalanian population databases. BMC Public Health. 2009; 25;9:202]. That article indicated that “The efficiency index (EI) will be established as the ratio: number of visits/expected visits for all patients (indirect standardization). An RI or EI value equal to one signifies equal complexity or efficiency to the standard (year 2008), whereas an EI < 1 symbolizes greater efficiency (inverse relation). “

In reference to the risk index (or morbidity burden), this has been used in other international publications. For example: “We calculated the risk index, also known as burden or index-cases, to relate the comorbidity of patients from the average of expected visits in 2007 (adjusted for age, sex and episodes) and those recorded in 2006 (by comparison method of indirect standardization)” [Sicras-Mainar A et al. Adjusted clinical groups use at a Spanish primary care center: a retrospective, population-based study. Rev Panam Salud Publica. 2010;27(1):49-55]. Another reference by the same first author also uses the Risk Index [Sicras-Mainar A et al. The relationship between effectiveness and costs measured by a risk-adjusted case-mix system: multicentre study of
Catalonian population databases. BMC Public Health. 2009; 25;9:202]: “The risk index (RI) reflects the complexity of the diagnoses attended by a centre with respect to the standard (case-mix).”

Construction of the synthetic index (SI) was based on 20 clinical indicators that are part of the standard health service contract with providers of healthcare services, attached to this reply [19. Technical report. Purchase of services 2007. Contracts for primary health care teams. Barcelona: Catalan Health Department; 2007]. These are the same consensus indicators used by CatSalut (the Catalan Health Service system) to assess and compare performance and patient profiles of its health care centres. We have added a paragraph to the revised manuscript (included in our response to the reviewer’s request that we “elaborate more on the sensitivity and specificity of the SI”, below).

2. The health care delivery system as mentioned and compared in the manuscript is also unique and specific to the local health care system. The authors redefine and explain the PHCs and BCTs and their relationship in the revised manuscript.

Thank you. We agree that the revision requested in the initial review will be useful to the reader.

3. A few typos have been corrected but I still can find one typo on page 6 that the ICD-9-CM was miscoding as ICD-9-MC.

We have corrected this mistake to ICD-9-CM and again searched the text for this specific error.

For the revised manuscript, I still have a few concerns for the authors of this great study:
1. Fewer parameters for comparison could be considered. There are EI, RI, and SI parameters and the EI can be further divided to three parts: cost, episode, and visit EI. Although the authors add a table 5 in explanation of the relationship between these five parameters, I am afraid that the majority of readers of the Journal are hard to interpret them in a simple way. Perhaps the authors can consider to deselect one, for example, the visit or episode EI. The authors used a lot of space to explain the differences and relationship between them both in Methods and Discussion sections. From my understanding, parts of the relationship between episode and visit could be explained in the Risk Index by their working definitions.

Thank you for your comments. We have eliminated the “EI visit” row from table 5.

2. I still concern using the Synthetic Index (SI) representing the effectiveness or
outcomes of the PHC and BCTs. Although the authors provide one Technical Report for reference (reference #19), I could not access and read it. Moreover, the validity and reliability of using a proxy index based on 20 indicators has not been studied yet. When we look at the Table 4, the SI of all 12 PHCs are quite similar to each other (except PHC7). The authors had better elaborate more on the sensitivity and specificity of the SI to the effectiveness of PHCs.

Thank you for this comment. We have added the following paragraph to the revised manuscript:

The synthetic index (SI) was obtained from a selection of 20 primary care process and outcomes indicators developed by CatSalut (Catalan Health Service). Originally obtained from the literature, the indicators were subsequently validated by an expert committee. These indicators reflect current standards for procedures related to primary and secondary prevention, diagnosis, treatment and patient monitoring (Table I). They address selected health objectives defined in the management contract between CatSalut and agencies providing primary healthcare services to the Catalan population [19]. For the implementation criteria, feasibility in the clinical setting was taken into account. The scores for indicators obtained from EHRs range from 1 to 100, reflecting a range from the lowest to the highest effectiveness.

We attach a complete copy of reference 19 for the reviewer’s convenience.

3. Given the homogeneity in PHC’s complexity (RI), effectiveness (SI), and three efficiency index (EIs), it is not surprising that the association between ACG-based efficiency measures and effectiveness parameters are weak (less than 0.5) though statistically significant. Again, the validity and sensitivity of these proxy variables are crucial.

We are not aware of any other study that has incorporated a measure of effectiveness into the internationally used ACG system. Therefore, we do not know whether the relationship is weak to moderate only at the statistical level or also at the clinical level. We do not have a measurement of any multiple correlation coefficient for reference or comparison. On the other hand, due to the homogeneity of the centres from which data were obtained for this study we are able to establish that these are centres with a similar complexity (or morbidity burden), reflected in the risk index – and with similar results in the efficiency index, based on costs, episodes, and visits. The fact that centres with such homogeneity yield differences in the synthetic index devised from our standard indicators in Catalonia indicates that despite their similarities they achieve different results in terms of their effectiveness. Therefore, it appears that some centres are more effective than others with equal resources.
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: Yes, and I have assessed the statistics in my report.
Declaration of competing interests: 'I declare that I have no competing interests.'