

Medición de la calidad de un proceso de codificación secundaria

García Martí Sebastián ^a, Arana Mónica ^b, Bazzolo Eleonora ^b, Salonia Sebastián ^b, Lopez Osornio Alejandro ^c, Langlois Esteban ^d, Luna Daniel ^e, González Bernaldo de Quirós Fernán ^e

^a Beca de Gerenciamiento Médico, ^b Área de Codificación, ^c Residencia de Informática Médica, ^d Gerencia Médica Plan de Salud, ^e Departamento de Información Hospitalaria, Hospital Italiano de Buenos Aires Argentina

Resumen

Este trabajo tiene como objetivo evaluar la calidad de la codificación secundaria en una lista de problemas generada a partir de texto libre escrito por los médicos, en el marco de uso de la Historia Clínica Electrónica del Plan de Salud, Hospital Italiano Buenos Aires. La medida de calidad a evaluar es la variabilidad entre codificadores bajo el principio de que dos textos libres iguales no pueden tener diferente código.

Se calcula un indicador de variabilidad y se propone una metodología de análisis de bases de datos codificadas secundariamente.

Palabras claves

Calidad bases datos, codificación secundaria, normalización texto libre.

Introducción

Los datos clínicos reales de los pacientes son la piedra fundamental en la que se basa la toma de decisiones en medicina, tanto en ámbitos asistenciales, de investigación o de gestión en salud. Por este motivo todos los sistemas de salud invierten gran cantidad de recursos en mantener bases de datos actualizadas que registran los eventos de la población bajo su cuidado.

Estas bases de datos pueden generarse automáticamente a partir de un Sistema de Información Hospitalario con un sistema de Historia Clínica Electrónica o manualmente con la transcripción de formularios llenados a mano. Para que estos datos sean pasibles de un análisis estadístico - informático es necesario un proceso de codificación en el cual se definen y estandarizan cada una de las entidades a medir mediante un vocabulario controlado, es decir que por ejemplo se transforman todas las diferentes formas de llamar a una infección pulmonar en el código correspondiente a 'Infección Pulmonar' en la clasificación Internacional de Enfermedades de la OMS. Esta *codificación* de los datos permite su agrupación y comparación estadística.

Los procesos de generación de las bases de datos tienen características que pueden hacer que el resultado final no refleje exactamente la realidad. Los sistemas de codificación que exigen al médico asistencial a elegir y asignar un código a un diagnóstico (Codificación Primaria), representan una sobrecarga laboral del mismo y son susceptibles de errores por falta de codificación. Cuando los médicos solo describen el diagnóstico con un texto narrativo y luego un especialista se dedica a codificarlo (Codificación Secundaria), hay una 'brecha' entre el diagnóstico y la codificación que puede llevar a una codificación errónea es decir que el código elegido por el codificador no represente la situación clínica del paciente. Para evitar esto y no tomar decisiones basándose en datos erróneos es necesario un estricto control de calidad de las bases de datos.

Hoy en día las formas de evaluación planteadas de las bases de datos, generalmente se basan en evaluar cuanto acuerda el contenido de las mismas con los eventos en salud que realmente se hayan producido en la población. Para esto se proponen comparaciones con elementos externos a la base como ser encuestas a los pacientes o el registro de papel, para después mediante el cálculo de la *completitud* (sensibilidad) y la *exactitud* (valor predictivo positivo) evaluar el grado de acuerdo entre el registro en la base de datos y la realidad. Cualquiera de estos dos abordajes implica tener un comparador externo de la base de datos para realizar la evaluación.

Por otro lado hay ciertas medidas de correlación interna de la base de datos que nos pueden aportar información sobre la validez de la misma sin utilizar comparaciones externas, lo que facilita la

evaluación siendo un método más rápido y menos costoso. Estas se basan en reglas que definen inequívocamente inconsistencias en los datos, como por ejemplo un varón no puede tener codificado un Papanicolau, así como una mujer no puede tener registrado un problema prostático.

En las bases de datos generadas mediante codificación secundaria se nos plantea un paso intermedio en la evaluación del acuerdo entre los códigos de la base de datos y los eventos en salud de la población, que es el de acuerdo entre el texto libre y los códigos.

En estas bases de datos con codificación secundaria hay principalmente dos fuentes de error, el defecto de *completitud* cuando no se registra en texto libre el evento ocurrido en la realidad y el defecto de *exactitud* cuando el texto libre es asignado a un código erróneo de la nomenclatura. Este proceso de codificación secundaria habitualmente se realiza en forma manual lo que conlleva un grado de error previsible.

En este trabajo se plantea una estrategia de evaluación de calidad del proceso de codificación secundaria, centrándose más que nada en la variabilidad de la misma. Se fundamenta en un análisis de la concordancia entre el texto libre y los códigos asignados basándose en que utilizando una clasificación sin ambigüedad dos descripciones en texto libre iguales no pueden tener diferentes códigos.

Objetivo

Desarrollar un índice de evaluación de la calidad de un proceso de codificación secundaria.

Material y Métodos

La base de datos utilizada para el análisis, es la base de la lista de problemas de los pacientes ambulatorios del Hospital Italiano de Buenos Aires, que es una prepaga de 90.000 afiliados que funciona inserta en el contexto de un Hospital Universitario en el ámbito de la Capital Federal, Buenos Aires.

La lista de problemas es una serie de diagnósticos que representan los problemas de salud de los pacientes. Para la generación de la base de datos, se cuenta con una historia clínica electrónica con terminales en cada consultorio que son usadas directamente por los profesionales de salud y de esta forma se genera un registro en texto libre que es codificado secundariamente por codificadores entrenados, mediante la asignación de códigos de la Clasificación Internacional de Atención Primaria.

La Clasificación Internacional de Atención Primaria (CIAP) es un sistema de clasificación compuesto por 800 rúbricas repartidas en 17 capítulos, de muy poca granularidad para ser utilizado como terminología de referencia pero especialmente diseñado para reflejar la realidad de los diagnósticos más frecuentes en Atención Primaria.

El registro electrónico genera más allá de la lista de problemas, la prescripción de fármacos, las interconsultas y los resultados de laboratorio, aunque a los fines de este trabajo solo se analizará la lista de problemas.

La base de datos contiene 524.248 diagnósticos con el texto narrativo ingresado por los médicos y el código CIAP asignado por los codificadores.

Estos 524.248 diagnósticos incluyen 104.354 textos repetidos, es decir que los médicos escribieron exactamente igual cada uno de los diagnósticos.

Distintos textos narrativos que describen la misma situación clínica pueden diferir solo en unos pocos caracteres que no cambien el sentido del diagnóstico, como puede ser una "S" final, un guión ("-") o barra ("/"), o simplemente dos palabras separadas por un espaciado doble en lugar de simple. Estos son identificados como diferentes en un análisis con un sistema de computación, así que sometimos todos los textos libres a un proceso de normalización en que quitamos los errores más comunes que pueden hacer diferente un texto sin cambiar su sentido.

Luego de este proceso quedaron 88.065 problemas diferentes, que a partir de ahora llamaremos

grupos de diagnósticos. La suma de todos los diagnósticos incluidos en estos grupos son los 524.248 diagnósticos iniciales.

Analizando estos 88.065 grupos vemos que 68.663 están compuestos por solo 1 diagnóstico, es decir que el texto narrativo ingresado por el médico nunca se repitió. En un método de evaluación interna como el que proponemos estos casos de una sola aparición no pueden ser comparados con nada para suponer su correctitud, así que son excluidos de todo el análisis posterior, ya que no se puede evaluar variabilidad de un texto libre que esta representado una única vez.

Los 19.402 grupos restantes contienen más de un diagnóstico cada uno, codificado manualmente con CIAP por los codificadores. No todos los grupos tienen un 100% de consistencia en la codificación debido a los errores inherentes a la codificación manual. Es decir que si por ejemplo los médicos escribieron "Neumonía" en 100 oportunidades, 85 fue codificada con el código CIAP "N20", 5 con el "N22" y 10 con el "P88". Nuestra medición de calidad se basa en que hay un solo código correcto para cada grupo y que indudablemente el resto pueden considerarse erróneos. Definimos como correcto el código asignado en la mayoría de las veces, requiriendo que la porción del grupo codificada con este código se mayor al 50%. Esta estrategia pasa por alto errores sistemáticos, en los que los codificadores asignan erróneamente los códigos en la mayoría de las oportunidades. Este error se valoró en un estudio previo en el que se evidenció una concordancia de los codificadores con un panel de expertos de un Kappa de 0,81 , un valor satisfactorio para nuestros objetivos.

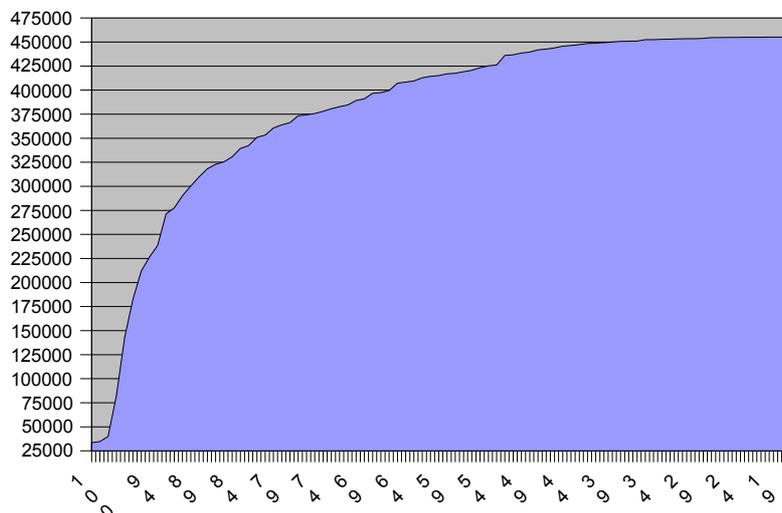


Figura 1: Cantidad de problemas (eje Y) según el grado de acuerdo en el grupo (eje X). Vemos que pocos tienen 100% de acuerdo pero muchos llegan al 90%. El criterio utilizado para la elección del código presuntamente correcto fue de la mayoría, es decir más del 50% de acuerdo en un código.

Los diagnósticos que se encuentran en grupos donde no existe un código que supere un 50% de acuerdo son considerados como erróneos. Lo mismo que los diagnósticos que se encuentran en grupos donde hay un código predominante, pero se encuentren en las porciones minoritarias.

Solo se consideran presumiblemente correctos los diagnósticos que aparecen más de una vez y fueron codificados con el código que tiene más del 50% de los códigos incluidos en su grupo.

Resultados

En los 19.402 grupos con más de una aparición hay 397.246 diagnósticos codificados con el código predominante (más del 50% del grupo con el mismo código), estos son considerados correctos por este índice y estrategia de control de calidad diferente.

Por otro lado en estos mismo grupos hay 58.239 diagnósticos con un código diferente al código predominante o incluidos en grupos sin predominancia mayor al 50%. Todos son considerados como erróneos.

El índice de Variabilidad se calcula como la proporción que representan los diagnósticos

“presumiblemente” codificados en forma errónea sobre el total de los diagnósticos, excluyendo aquellos que aparecen una única vez.

En nuestra base de datos en el índice de variabilidad de la codificación fue del 12,78%.

Discusión

El proceso de evaluación de calidad de bases de datos, al ser las bases de dimensiones considerables se hace dificultoso y la comparación con patrones externos costosa, por otro lado es la única forma de evaluar la completitud de la misma.

En trabajos anteriores vemos que en contadas ocasiones se utilizaron patrones externos como comparadores.

Al evaluar la exactitud de la base, la variabilidad de la codificación es uno de los factores que influyen en la misma, sabiendo que si dos textos libres tienen diferente código ya alguno de los dos no es exacto, pero no sabemos cual, este es el punto donde el comparador externo se hace indispensable nuevamente.

Nosotros proponemos un índice de *supuesta* exactitud debido a que para completar el trabajo habría que muestrear en estos grupos y ver que en los grupos en los cuales alguno de los diagnósticos era mayor al 50%, este código coincidiera realmente con el texto libre del grupo y de esta forma podríamos estar evaluando la exactitud y no solamente la variabilidad que es uno de sus componentes.

Las ventajas de la estrategia planteada son su rapidez y bajo costo, tiene como desventaja que no evalúa la completitud de la base a comparación de las que utilizan comparadores externos, y en cuanto a la exactitud nos da una idea aproximada pero no la define detalladamente.

Según los resultados podemos dividir todos los diagnósticos en la base de datos en 3 grupos:

Los diagnósticos con una codificación “**presumiblemente correcta**”, que deben ser validados con la recodificación de una muestra al azar para buscar errores sistemáticos y de esta forma evaluar la exactitud de los mismos.

Los diagnósticos “**presumiblemente incorrectos**” al estar codificados con un código que no tiene predominancia. Algunos de estos pueden ser re-codificados automáticamente luego de descartar la posibilidad de un error sistemático de acuerdo al código predominante. Los que no tengan código predominante deben ser revisados término por término para que los codificadores lleguen a un acuerdo en el código a elegir para ese término.

Por último los **diagnósticos “presumiblemente incorrectos”** por no haberse repetido nunca y no poder hacer una estimación automática. Estos deben revisarse uno por uno, pero mejoras en la estrategia de normalización pueden llegar a incluirlos en otros grupos con más repeticiones y luego incluirlos en uno de los grupos anteriores. A su vez capacitación de los médicos en la forma de expresar los problemas en la historia clínica pueden hacer disminuir el número de estos diagnósticos.

Conclusión

El sistema propuesto permite realizar una inferencia de la exactitud de una base de datos utilizando medidas internas que evalúan la variabilidad, representando una medida rápida y eficiente. basándose en esta herramienta es posible organizar las tareas de vigilancia y mejoramiento de la calidad (exactitud) de la base de datos. Este índice no evalúa la Completitud de la base de datos.

Bibliografía

Brennan PF, S.W., Assessing Data Quality: From Concordance, through Correctness and Completeness, to Valid Manipulable Representations. J Am Med Inform Assoc., 2000. 7(1): p. 106-7.

Hassey A, G.D., Wilson A., A survey of validity and utility of electronic patient records in a general practice. *BMJ*, 2001. 322: p. 1401-5.

Hogan WR, W.M., Accuracy of data in computer based patients records. *J Am Med Inform Assoc.*, 1997. 4(5): p. 342-55.

Horbar JD, L.K., An assesment of data quality in the Vermont-Oxford Trials Network database. *Controlled Clin Trials*, 1995. 16: p. 51-61.

Iezzoni LI, Assessing Quality Using Administrative Data. *Ann Intern Med.*, 1997. 127: p. 666-74.

Luna, D., F.G. Bernaldo de Quiros, L. Garfi, E. Soriano and M. O'Flaherty, Reliability of secondary central coding of medical problems in primary care by non medical coders, using the International Classification of Primary Care (ICPC). *Medinfo*, 2001. 10(Pt 2): p. 300.

McDonald, Quality measures and electronic medical systems. *JAMA*, 1999. 282(12): p. 1181-2.

Schneider EC, R.V., Courte-Wienecke S, Eddy DM, Sennet C., Enhancing performance measurement: NCQA's road map for a health information framework. *JAMA*, 1999. 282(12): p. 1184-90.

Schpilberg, M., A. Lopez Osornio, F. Pedernera, D. Luna, G. Rodriguez, E. Soriano, A. Gomez, M. Martinez, M. Marchetti and F.G. Bernaldo de Quiros. Sensibilidad de la historia clínica electrónica para el registro de fármacos. in 5to Simposio de Informática en Salud - 31 JAIIO. 2002. Santa Fe, Argentina: SADIO.

Schpilberg, M., D. Luna, G. Rodriguez, S. Figar, E. Soriano, A. Gomez, M. Martinez and F.G. Bernaldo de Quiros. Sensibilidad y Especificidad de la Historia Clínica Electrónica para registrar patología crónica. in 5to Simposio de Informática en Salud - 31 JAIIO. 2002. Santa Fe, Argentina: SADIO.