MEETING OF THE WHO COLLABORATING CENTRES FOR THE FAMILY OF INTERNATIONAL CLASSIFICATIONS

10 – 16 October 2009
Seoul, Republic of Korea

D037p

Assessment of automated coding of mortality in Mexico. 2007

Authors: Solis J, Ortega A, Suarez M.
INEGI and CEMECE: México

Abstract After some years of work, the National Office of Statistics of Mexico (INEGI) implemented in 2007 the System of Automated Coding of the Underlying Cause of the Death (SCACD). This change involves benefits that need to be supported, to ensure the operation of the system to provide a high degree of reliability and quality of mortality statistics. This paper shows the methodology to assess the proportion of correct coding (records) by the system, this results are based on statistical methods and the support of experts in death coding in Mexico.

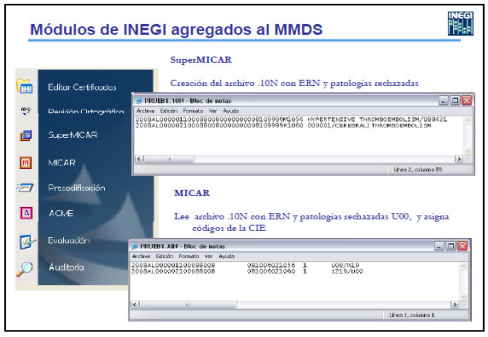
Introduction

The selection of underlying cause of death manually means a complex problem, it requires a high degree of specialty for the application of rules and coding procedures, contained in the ICD-10.

It requires a long time to become an expert in order to collaborate to the generation of high quality mortality statistics.

To solve this problem, the INEGI has been adapted to Spanish, the Mortality Medical Data System (MMDS) developed by CDC/NCHS of USA. Such adaptation includes new modules like verification and audit of coding.

SCACD modules



This cartel presents the methodology to assess the proportion of correct coding (records) by the SCACD.

Methods & Materials

370,662 cases coded during year 2007 with SCACD were used to obtain a probabilistic sample.

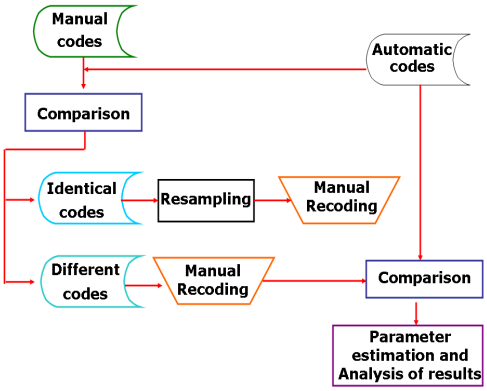
The sample design, was under a sampling stratified by group, with simple random selection without replacement within each stratum.

It was calculated the size of the sample for the correct encoding, with an absolute error of estimation does not exceed 1% and with 95% confidence level for the total and for each stratum.

Further information:
Javier Francisco Solis Guerrero
Javier.solis@inegi.org.mx

Methods & Materials

The matching process



It was selected a sample of 12,681 records coded by the system, recoded again manually by a team of 12 expert coders. They didn't know the code assigned by the system. At the end, both codes were compared.

Those that were equal were resample again to quantify the mistake made by, the system and the coder in a fortuitous mistake. Different cases were reviewed and coded by the team of experts.

These cases were assigned to different coders and they didn't know the first manual code, or the code from the automated system.

The new comparison covered three types of independent coding; two of the three codes matching, determined the success or failure of automatic coding.

The few cases where the three codes where different, were reviewed by the INEGI encoder in Chief, to determine the correct code to know if the systems code was right or wrong.

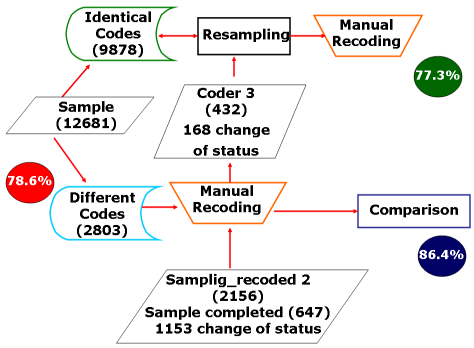
The estimate of the proportion and its confidence interval was calculated by:

$$\hat{p}_{st} = \frac{1}{N}\sum_{h=1}^{H} N_h \hat{p}_h$$

$$IC = \hat{p}_{st} \pm \sqrt{\left(\frac{z_{\alpha/2}}{N^2 n}\left(\sum_{h=1}^{H} N_h \sqrt{p_h q_h}\right)^2 - \frac{z_{\alpha/2}}{N^2}\sum_{h=1}^{H} N_h p_h q_h\right)}$$

Results

Results of the comparison



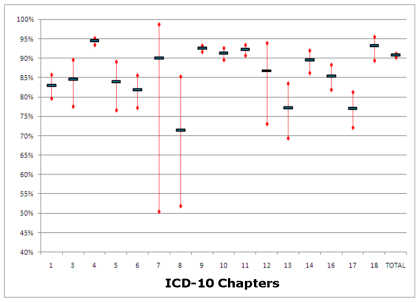
At the end of this process, with the new system (SCACD) version the percentage of correct coding was 90.8%.

Proportions and confidence by chapter

CAPÍTULO	DESCRIPCIÓN	Sistema Nuevo		Intervalo de confianza (95%)	
		PROPORCION	E.ESTANDAR	Mínimo	Máximo
1	CIERTAS ENFERMEDADES INFECCIOSAS Y PARASITARI	82.95%	1.02%	75.74%	90.70%
3	ENFERMEDADES DE LA SANGRE Y DE LOS ORGANOS H	84.51%	3.05%	77.56%	89.99%
4 *	ENFERMEDADES ENDOCRINAS, NUTRICIONALES Y MET	94.45%	0.46%	93.47%	95.29%
5	TRASTORNOS MENTALES Y DEL COMPORTAMIENTO	83.82%	3.17%	76.62%	89.12%
6	ENFERMEDADES DEL SISTEMA NERVIOSO	81.85%	2.14%	77.27%	85.67%
7	ENFERMEDADES DEL OJO Y SUS ANEXOS	98.00%	10.00%	56.48%	98.76%
8	ENFERMEDADES DEL OIDO Y DE LA APOFISIS MASTOID	71.45%	8.69%	52.08%	85.21%
9 *	ENFERMEDADES DEL SISTEMA CIRCULATORIO	92.58%	0.41%	91.73%	93.35%
10 *	ENFERMEDADES DEL SISTEMA RESPIRATORIO	91.26%	0.75%	89.67%	92.63%
11 *	ENFERMEDADES DEL SISTEMA DIGESTIVO	92.21%	0.67%	90.75%	93.43%
12	ENFERMEDADES DE LA PIEL Y DEL TEJIDO SUBCUTAN	86.67%	5.12%	73.16%	93.94%
13	ENFERMEDADES DEL SISTEMA OSTEOMUSCULAR Y DE	77.21%	3.61%	69.38%	83.51%
14	ENFERMEDADES DEL SISTEMA GENITOURINARIO	84.51%	1.48%	81.73%	87.08%
15	CIERTAS AFECCIONES ORIGINADAS EN EL PERIODO PE	85.36%	1.04%	83.26%	87.31%
17	MALFORMACIONES CONGENITAS, DEFORMIDADES Y A	77.02%	2.85%	72.10%	81.80%
18	SINTOMAS, SIGNOS Y HALLAZGOS ANORMALES CLINIC	91.47%	1.52%	89.58%	93.90%
TOTALES		90.753%	0.252%	90.247%	91.235%

* Capítulos con mayor frecuencia

Confidence intervals



Conclusions

The correct coding by the automated coding system (SCACD) was 90.8 %.

The chapters with the largest number of correct cases reached more than 91 % with the exception of chapter 19 while chapter 8 only achieve 72.4% of correct cases.

It is possible to implement this methodology every year to evaluate the statistics of deaths, in order to validate automatic coding system and therefore increase the quality of statistics generation process.