

**DEVELOPMENT OF A CODING AND CROSSWALK TOOL
FOR OCCUPATIONS AND INDUSTRIES**

Rémen T¹, Richardson L¹, Pilorget C^{2,3}, Palmer G⁴, Siemiatycki J¹, Lavoué J¹

¹ University of Montreal Hospital Research Center (CRCHUM), Montreal, Canada

² The French Public Health Agency, Saint Maurice, France

³ Univ Lyon, University of Claude Bernard Lyon1, Ifsttar, UMRESTTE, UMR T_9405, F- 69373, Lyon, France.

⁴ French Center for Research and Development in Medical Informatics (CREDIM), ISPED, Bordeaux, France

Table S1: Proportion of jobs in a database of occupational histories (31,673 jobs) for which the CAPS-Canada crosswalk assistant allowed linkage between pairs of industry classification systems

Given classification	Target classification	Proportion of jobs linkable with the full crosswalk assistant		Proportion of jobs linkable with only “official” links	
		< 10 codes proposed ^α	direct link to one job ^β	< 10 jobs	1 job
ISIC-1971	NAICS-2012	78.7%	17.3%	18.2%	2.3%
ISIC-1971	SIC-1980	71.9%	21.0%	19.3%	1.6%
NAICS-2012	SIC-1980	100.0%	74.4%	94.4%	47.7%
NAICS-2012	ISIC-1971	99.4%	90.4%	87.2%	35.5%
SIC-1980	ISIC-1971	99.0%	90.9%	77.8%	32.6%
SIC-1980	NAICS-2012	99.6%	80.0%	93.0%	40.9%

For each possible pair of starting / target classifications was evaluated for what proportion of jobs in our database the crosswalk from the starting to the target classification would lead to:

^α “< 10 codes proposed” = useful information provided

^β “direct link to 1 code” = univocal links

Table S2: Inter-rater reliability in industry coding between the two coders for each classification used

	Resolution level (number of Digit)	Number of corresponding codes at the resolution level	Inter-rater reliability in job coding [95% CI] - All approaches combined
ISIC (1971)	4	159	78.5% [75.9% -81.1%]
	3	71	81.1% [78.6% -83.6%]
SIC (1980)	4	860	64.1% [61.1% -67.1%]
	3	318	71.3% [68.4% -74.2%]
SCIAN (2012)	6	922	63.2% [60.2% -66.2%]
	5	711	67.3% [64.3% -70.3%]
	4	323	71.5% [68.7% -74.3%]
	3	102	78.2% [75.6% -80.8%]

Table S3: Correspondences between industry classifications based on “official” crosswalks

		Target classification						
		6-digit NAICS-2012		4-digit SIC-1980		4-digit ISIC-1971		
		$N_{\text{Codes}}^{\alpha}$	$N_{\text{Tables}}^{\beta}$	Median [5 th - 95 th percentile] ^{γ}	$N_{\text{Tables}}^{\beta}$	Median [5 th - 95 th percentile] ^{γ}	$N_{\text{Tables}}^{\beta}$	Median [5 th - 95 th percentile] ^{γ}
Starting classification	6-digit NAICS-2012	922			2	1 [1 - 5]	2	2 [1 - 23]
	4-digit SIC-1980	860	2	1 [1 - 6]			2	2 [1 - 37]
	4-digit ISIC-1971	158	2	20 [1.9 - 99.6]	2	26 [2 - 109.5]		

^{α} Number of different codes for the classification

^{β} Number of intermediate tables used between the starting and target classifications for “official” crosswalks

^{γ} Median and values of the 5th and 95th percentile of the distribution of numbers of possible codes in the target classification linked with each code in the starting classification

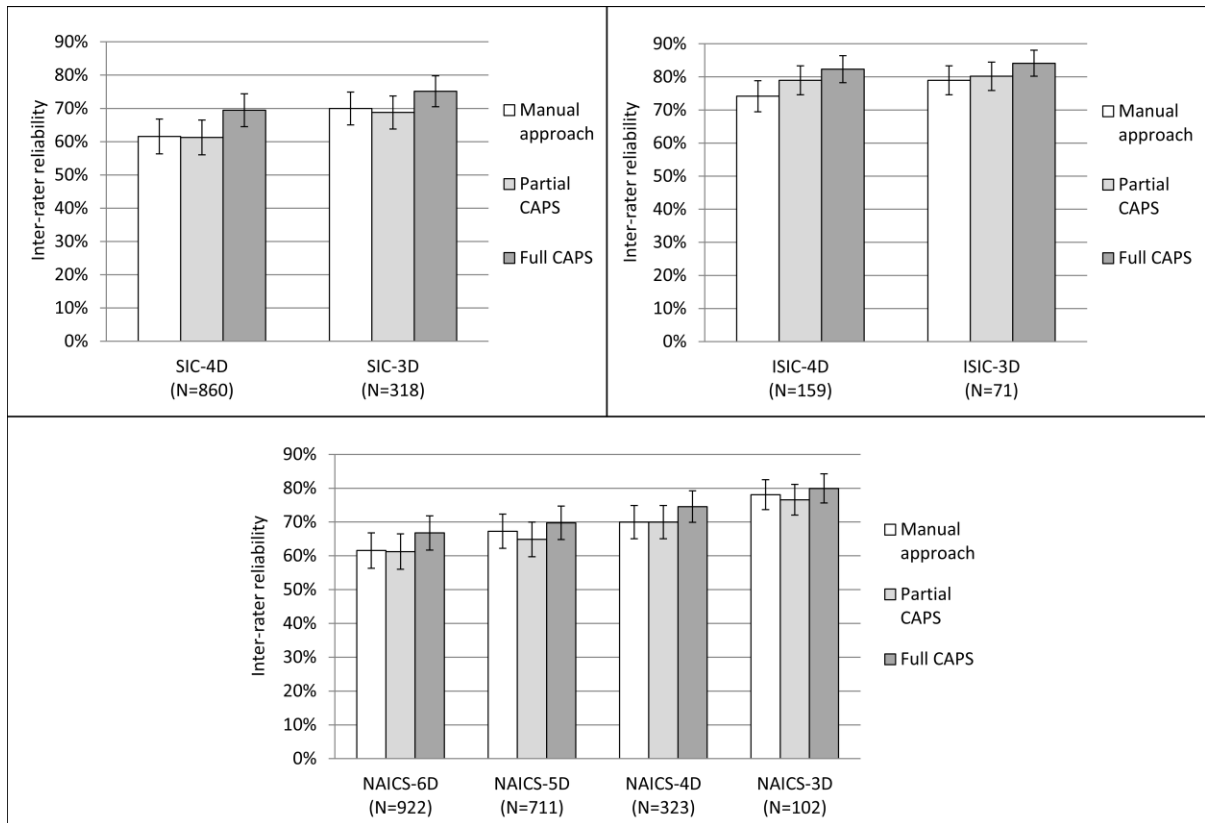
Table S4: Relative contribution of the different types of link covered by the crosswalk assistant

		Proportion of jobs in a database of occupational histories (31,673 jobs) for which crosswalks provide useful information							
Origin system	Target system	≤ 10 jobs proposed				Direct link to 1 code			
		All links combined ^β	By type of link (exclusively) ^α			All links combined ^β	By type of link (exclusively) ^α		
			"Expert"	"Empirical"	"Official"		"Expert"	"Empirical"	"Official"
ISIC-1971	NAICS-2012	78.7%	0%	78.5%	0.1%	17.3%	0%	17.2%	0.0%
ISIC-1971	SIC-1980	71.9%	0%	71.8%	0.1%	21.0%	0%	21.0%	0.0%
NAICS-2012	SIC-1980	100.0%	0%	95.3%	4.7%	74.4%	0%	71.3%	3.1%
NAICS-2012	ISIC-1971	99.4%	0%	95.3%	4.1%	90.4%	0%	87.1%	3.4%
SIC-1980	ISIC-1971	99.0%	2.4%	94.0%	2.6%	90.9%	2.4%	86.8%	1.8%
SIC-1980	NAICS-2012	99.6%	9.6%	86.8%	3.2%	80.0%	9.6%	68.5%	1.9%

^α The category of links was ranked from the most to the less informative. When a link is available for a starting code, links at the lower levels were not sought

^β The sum can vary due to rounding

Figure S1: Inter-rater reliability (and 95% confidence interval) in industry coding according to the approach used



Manual approach: manual coding based on published version of the classifications - Partial CAPS: CAPS coding assistant + “official” crosswalk assistant - Full CAPS: the full CAPS-Canada tool including all crosswalk types

Classifications and versions retained: SIC-1980, ISIC-1971 and NAICS-2012

The figures following the classification indicate the number of digits kept and the number in brackets refers to the number of different codes for the classification/resolution level combination