

Karine Dupriez, Ignazio Cassis, John-Paul Vader

Institut universitaire de médecine sociale et préventive, Lausanne

Utilisation de la classification internationale des maladies dans l'assurance-invalidité en Suisse: Étude de reproductibilité

Résumé

La Classification Internationale des Maladies, dixième révision (CIM-10) a été introduite dans un échantillon de cas rencontrés dans l'assurance-invalidité (AI), à titre d'essai, pour le codage des diagnostics, en substitution aux codes actuels, propres à l'AI et jugés inadéquats à la description de l'invalidité. Cet essai concerne uniquement des cas examinés aux Centres d'observation médicale de l'AI (COMAI). Le but de cette étude est d'évaluer, d'une part, la fiabilité d'utilisation de ces codes – en étudiant la reproductibilité de codage entre le Centre d'observation et d'évaluation (CECO) et les COMAI – et, d'autre part, la faisabilité du codage par des non-médecins. Les diagnostics ont été divisés en niveaux, fonction de leur précision, respectant la structure de la CIM-10. Il s'agit des chapitres, groupes, catégories et sous-catégories. La concordance entre les 2 centres codeurs varie de 78% au niveau du chapitre à 29% au niveau de précision le plus élevé. Cette concordance est améliorée en considérant les 3 diagnostics principaux dans la comparaison, atteignant alors 91% pour le niveau du chapitre. Il existe des différences de concordance inter-juge selon le COMAI de référence. Les résultats de reproductibilité (codage-recodage) du médecin sont meilleurs par rapport aux non-médecins, notamment une économiste-statisticienne et une secrétaire médicale. La concordance de codage pourrait être améliorée, notamment en formant les codeurs, en proposant des outils d'aide au codage, en adoptant une définition précise pour ce qu'est le premier diagnostic principal et en trouvant un consensus sur les discordances systématiques. Le niveau de précision idéal à atteindre ne doit pas être trop élevé pour ne pas fournir des diagnostics discordants, tout en étant suffisamment précis pour fournir une bonne description des pathologies rencontrées.

Pour que la politique sanitaire mène à bien ses tâches de planification, d'évaluation et de contrôle des services de santé^{1, 2}, elle doit s'appuyer sur un système d'infor-

mation performant. Ces informations doivent répondre, entre autres, à des critères de précision et de fiabilité. La statistique médicale – dont les diagnostic et leurs codes –

fait partie de ces informations indispensables au fonctionnement de la politique sanitaire et, à ce titre, doit satisfaire à ces critères. L'assurance-invalidité (AI) emploie depuis les années 1960 des codes créés par et pour elle, qui répondaient originellement à un but comptable. L'utilisation de ces codes à des fins épidémiologiques ou de gestion s'est avérée difficile, voire impossible, à cause de leur hétérogénéité et de leur imprécision, amenant même la question posée dans une étude récente: «les codes AI décrivent-ils l'invalidité?»³.

La même étude récente a conclu qu'une classification des diagnostics – causes de l'invalidité – devrait être complétée par l'utilisation d'une classification des conséquences de l'invalidité. La Classification Internationale des Maladies, 10ème révision (CIM-10) avait été proposée pour la description des causes et la Classification internationale des handicaps (ICIDH) pour la description des conséquences de l'invalidité. Compte tenu de l'important travail qu'impliquerait pour l'AI l'introduction, sans nouvelle ressources, à la fois d'une nouvelle classification des diagnostics et d'une classification des handicaps, le Centre d'évaluation et de coordination des COMAI (CECO) a décidé d'introduire, à

titre d'essai, uniquement le CIM-10 pour le codage des diagnostics et ce, exclusivement pour les cas traités par les Centres d'observation médicale de l'assurance-invalidité (COMAI). Les avantages majeurs de cette classification, montrés ailleurs, sont la possibilité de comparer la mortalité et la morbidité à un niveau international et sur de longues périodes et la possibilité d'une grande précision et spécificité des données⁴. Disposant maintenant d'un outil de bonne qualité, il convient d'examiner si son utilisation répond aux critères précités. L'objectif principal de cette étude est d'évaluer la fiabilité d'utilisation de ces codes en examinant la reproductibilité inter-juges; un codage uniforme, pour le plus grand nombre de cas et à un niveau de précision élevé, étant la base nécessaire à la bonne qualité des données. Dans un second temps, l'hypothèse de faisabilité du codage par des non-médecins sera testée, ainsi que la reproductibilité intra-juge à deux temps. Ceci permettra de déterminer le profil des collaborateurs à qui confier cette tâche parmi ces différents professionnels.

Méthodes

Les cinq COMAI suisses examinent environ 1400 assurés sur les 60000 s'adressant à l'AI chaque année; ils s'occupent des cas dont la capacité de travail – composante principale pour le calcul du degré d'invalidité – est difficile à évaluer et qui nécessitent un examen médical pluridisciplinaire. A la suite de l'expertise, les médecins des COMAI attribuent, au maximum, 3 codes CIM-10 relatifs à la pathologie principale. Les dossiers d'expertise de ces assurés parviennent ensuite au CECO où les diagnostics sont codés une seconde fois, par un médecin, à partir de la description de la pathologie faite dans le dossier. Pour chaque cas, il existe donc deux codages effectués

à deux moments, à deux endroits et par deux codeurs différents.

Les diagnostics ont été divisés en niveaux, selon leur précision, respectant la structure de la CIM-10. Il s'agit des chapitres (exemple: J, maladie de l'appareil respiratoire), des groupes (exemple: J00–J06, affections aiguës des voies respiratoires supérieures), des catégories (exemple: J01, sinusite aiguë) et des sous-catégories (exemple: J01.1, sinusite frontale aiguë)⁵. Seul le premier diagnostic principal a été pris en considération, sauf pour le tableau 4 qui traite des 3 diagnostics principaux ensemble.

La reproductibilité inter-juges entre les COMAI et le CECO se base sur un échantillon de 1002 cas. Sur les 1395 cas examinés aux COMAI en 1994, 70 ne comportaient pas de dossiers d'expertise, rendant impossible le codage au CECO et 323 autres provenaient d'un COMAI utilisant encore l'ancienne classification. Soixante deux pour-cent de l'échantillon sont des hommes, la moyenne d'âge est de 45 ans.

La reproductibilité entre juges de formations différentes est examinée à partir d'un échantillon de 150 expertises, tirées aléatoirement parmi les 1002 valides de l'année 1994. Un seul diagnostic a été pris en compte, extrait du dossier COMAI par le juge 1. Ce dossier est un médecin possédant une année de pratique du codage avec la CIM-10. Le juge 2 est économiste-statisticien; il a une expérience professionnelle dans le domaine de l'assurance-invalidité mais pas de pratique du codage. Le juge 3 est secrétaire médicale sans connaissance particulière de l'AI, ni du codage. Pour évaluer la reproductibilité intra-juge, les 3 juges ont recodés 40 des 150 expertises à un mois d'intervalle. Pour cette dernière évaluation, un score moyen de reproductibilité a été établi en attribuant la valeur 1 pour un codage concordant au niveau du chapitre, 2 au niveau du groupe, 3 au niveau de la catégorie et 4 au niveau de la sous-catégorie.

Les test du chi-carré a été employé pour évaluer la significativité statistique des résultats entre les cinq COMAI. L'intervalle de confiance exacte a été calculé pour le pourcentage de concordance interjuge et le t-test utilisé pour examiner la différence du score moyen de reproductibilité entre des paires de juges. La concordance entre les deux sites de codage a été évaluée avec le coefficient kappa⁶.

Résultats

Les diagnostics rencontrés dans la population COMAI sont très spécifiques: deux chapitres représentent plus de $\frac{3}{4}$ des cas. Même au niveau d'agrégation le plus fin, les 5 codes les plus utilisés décrivent plus de 20% des cas (Tableau 1).

La concordance entre les 2 centres codeurs est de 78% au niveau du chapitre; au niveau de précision le plus élevé, elle n'est plus que de 29%. La concordance entre le codage du CECO et des COMAI A, B et E est meilleure que la moyenne, elle est moins bonne avec C et D (Tableau 2).

La reproductibilité du codage entre le CECO et les COMAI ne varie pas en fonction du sexe, de l'âge des patients, du trimestre ou de la longueur du séjour (résultats non présentés ici).

En se limitant aux chapitres V et XIII (Tableau 3), ceux qui sont les plus rencontrés, la concordance de 81% ($k = 0,70$, $p < 0,01$) au niveau du chapitre (diagonale) est importante⁶.

Si l'on considère les 3 diagnostics principaux dans la comparaison inter-codeurs, la concordance s'en trouve améliorée. En cherchant une concordance entre le premier diagnostic COMAI et les trois diagnostics CECO (et vice versa), elle peut atteindre 91% au niveau du chapitre et 36% au niveau de la sous-catégorie (Tableau 4).

La concordance entre juges, de formations différentes, varie de 89

Codage COMAI (n = 2162)				Codage CECO (n = 2170)			
	Codes	n	%		Codes	n	%
<i>Chapitres</i>							
1	XIII	913	42,2	XIII	948	43,7	
2	V	842	38,9	V	757	34,9	
3	XIX	100	4,6	XIX	145	6,7	
4	Autres	307	14,2	Autres	320	14,7	
<i>Categories/sous-categories</i>							
1	F45	123	5,7	M47.8	144	6,6	
2	M54	118	5,5	F45.4	125	5,8	
3	M47.8	71	3,3	M51.2	75	3,5	
4	F60	64	3,0	F45.2	65	3,0	
5	F32.1	55	2,5	M7.0	61	2,8	
6	Autres	1731	80,0	Autres	1700	78,3	

Chapitres:
 XIII (M00–M99) = maladies du système ostéo-articulaire, des muscles et du tissu conjonctif.
 V (F00–F99) = troubles mentaux et du comportement.
 XIX (S00–T98) = lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes.

Sous-catégories:
 F32.1 = épisode dépressif moyen.
 F45 = troubles somatoformes.
 F45.2 = trouble hypocondriaque.
 F45.4 = syndrome douloureux somatoforme persistant.
 F60 = troubles spécifiques de la personnalité.
 M47.8 = autres spondylarthroses.
 M51.2 = hernie d'un autre disque intervertébral [que cervical] précise.
 M54 = dorsalgies.
 M79.0 = rhumatisme, sans précision.

Tableau 1. Distribution des codes les plus fréquents, en considérant les trois diagnostics principaux, au niveau le moins détaillé (chapitres) et au niveau le plus détaillé (catégories ou sous-catégories).

Niveau de précision	COMAI										Total	
	A		B		C		D		E		n	% cumulé
	n	% cumulé	n	% cumulé	n	% cumulé	n	% cumulé	n	% cumulé		
Sous-catégorie	86	27	84	42	34	20	58	24	31	46	293	29
Catégorie	68	48	27	55	35	41	33	37	12	64	175	47
Groupe	70	71	40	75	27	57	52	58	13	84	202	67
Chapitre	32	80	15	83	20	71	41	75	0	84	108	78
Discordant	64		35		52		62		11		224	
Total	320	100	201	100	168	100	246	100	67	100	1002	100

Chi-carré = 63,4; p < 0,001.

Tableau 2. Concordance entre le codage COMAI et le codage CECO, selon le COMAI et le niveau de précision du premier code diagnostic. Le pourcentage cumulé de concordance CECO – COMAI est indiqué pour chaque niveau de précision.

	Chapitres	V	COMAI		Total
			XIII	Autre	
C	V	268	28	27	323
E	XIII	35	377	24	436
C	Autres	25	54	164	253
O					
	Total	328	459	215	1002

kappa = 0,70; p < 0,001.

Tableau 3. Concordance du codage des diagnostics selon la CIM-10 au niveau du chapitre entre les Centre d'observation médicale de l'AI (COMAI) et le Centre d'évaluation et de coordination (CECO).

à 98% au niveau des chapitres, jusqu'à 32 à 39% au niveau des sous-catégories (Tableau 5), semblable à la reproductibilité trouvée lorsqu'il s'agit de deux médecins qui codent. Considérant le médecin (juge 1) avec une année d'expérience en codant avec la CIM-10 comme l'étalon, on remarque que la secrétaire médicale (juge 3) semble coder de manière plus proche de ce critère que l'économiste-statisticienne (juge 2).

La reproductibilité intra-juge varie de 72 à 97%, selon le niveau de précision du code, pour le juge 1; elle est plus basse pour les juges 2 et 3, quel que soit le niveau

Niveau de précision	1er code COMAI comparé aux 1er, 2ème, 3ème codes CECO		1er code CECO comparé aux 1er, 2ème, 3ème codes COMAI	
	n	% cumulé	n	% cumulé
4 – Sous-catégories	364	36	339	34
3 – Catégories	223	59	216	55
2 – Groupes	218	80	233	79
1 – Chapitres	106	91	90	88
0 – Discordant	91		124	
Total	1002	100	1002	100

Tableau 4. Concordance entre le codage COMAI et le codage CECO, selon le niveau de précision du code diagnostic, en considérant les 3 premiers diagnostics principaux.

Niveau de précision	Juge 1–Juge 2		Juge 1–Juge 3	
	n	% cumulé (IC .95)	n	% cumulé (IC .95)
4 – Sous-catégories	48	32 (25; 40)	59	39 (31; 48)
3 – Catégories	29	51 (43; 60)	25	56 (48; 64)
2 – Groupes	36	75 (68; 82)	37	81 (73; 87)
1 – Chapitres	20	89 (82; 93)	26	98 (94; 99)
0 – Discordant	17		3	
Total	150	100	150	100

Tableau 5. Reproductibilité entre juges, de formation différente, du codage diagnostic utilisant l'ICD-10, selon le niveau de précision du code diagnostic (n = 150). Pour chaque niveau de précision le pourcentage cumulé de concordance est indiqué avec un intervalle de confiance (IC) 95%. Le juge 1 (médecin, codeur expérimenté) a été pris comme référence.

Niveau de précision	Juge 1		Juge 2		Juge 3	
	n	% cumulé (IC .95)	n	% cumulé (IC .95)	n	% cumulé (IC .95)
4 – Sous-catégories	29	73 (56; 85)	20	50 (34; 66)	21	53 (36; 68)
3 – Catégories	6	88 (73; 96)	4	60 (43; 75)	4	63 (46; 77)
2 – Groupes	4	98 (87; 100)	4	70 (53; 83)	5	75 (59; 87)
1 – Chapitres	–	98 (87; 100)	8	90 (76; 97)	9	98 (87; 100)
0 – Discordant	1		4			
Total	40	100	40	100	40	100

Tableau 6. Reproductibilité intra-juge du codage diagnostique utilisant l'ICD-10, à un mois d'intervalle, selon le niveau de précision du code diagnostique (n=40). Pour chaque niveau de précision le pourcentage cumulé de concordance est indiqué avec un intervalle de confiance (IC) 95%.

(Tableau 6). Lorsque cette reproductibilité est évaluée par un score moyen, le juge 1, avec un score moyen de 3,6, est significativement meilleur ($P < 0,01$) que les autres juges (score moyen 2,7 et 2,9 respectivement), qui ne se distinguent pas entre eux ($P = 0,6$).

Discussion

Une concordance insatisfaisante

Selon la spécialité étudiée, le niveau d'agrégation des codes, la rareté des diagnostics et le nombre de codeurs, les concordances rencontrées dans la littérature varient de 29 à 63% pour une grande précision (sous-catégories), de 41 à 87% pour une précision moyenne (catégories) et de 46 à 94% pour une faible précision (groupes)^{1,7-10}. Nos résultats se situent à l'intérieur de ces intervalles, bien que dans les tranches inférieures (voir Tableau 2). Cette concordance insuffisante, s'il n'a pas d'influence sur l'occurrence des diagnostics au niveau des chapitres, entraîne, par contre, des différences notables de statistiques au niveau des sous-catégories.

Hypothèses et propositions d'amélioration

A la décharge partielle de ces résultats, médiocres, il est à relever

qu'un des chapitres les plus rencontrés auprès de la population COMAI (39% des cas pour les 3 diagnostics principaux) est le chapitre V, troubles mentaux et du comportement, et qu'il a été montré ailleurs que ce chapitre présentait une variabilité de codage maximale, les désordres mentaux étant plus sujets à l'interprétation médicale^{9,11}. D'autre part, les discordances les plus rencontrées sont souvent le fait de diagnostics proches. Par exemple, des lombalgies non-spécifiques peuvent être codées dans les Maladies du système locomoteur (M54: Dorsalgies) dans un centre, mais dans les Troubles mentaux (F45: Troubles somatoformes) dans un autre. Les deux codes peuvent être tout aussi légitimes. Les divergences de cette sorte pourraient être éliminées, ou du moins diminuées, par des discussions de consensus puisqu'elles ne sont pas le fruit du hasard et conduisent à des désaccords systématiques.

Cette étude ne posait pas l'un des deux centres codeurs comme référence. Il est difficile d'évaluer lequel des deux codages est le plus «juste». Des sources d'«erreurs» peuvent exister des deux côtés. De la part des médecins exerçant aux COMAI, on peut imaginer un manque d'entraînement au codage et une méconnaissance des règles de codage ainsi qu'un manque d'intérêt et d'implication, contrairement

au CECO qui porte un véritable intérêt scientifique à ce domaine. Par contre, le codage de ce dernier s'étant fait de manière rétrospective, à partir des dossiers médicaux, on peut voir là une vraie faiblesse. Il a été montré que quand les praticiens ont à coder des informations ne relevant pas directement de leurs expertises, ils ont davantage recours à des codes non-spécifiques (0,9 ou s'arrêtant au 2ème chiffre)¹⁰. Ce n'est pas le cas ici, où ces codes non spécifiques sont utilisés par le CECO dans 22% des cas contre 41% aux COMAI. Ce résultat peut suggérer que le codage au CECO a été trop précis par rapport aux informations disponibles et qu'il résulte d'une extrapolation excessive.

Un problème de langue a pu gêner ce travail: aucun des centres n'a pu disposer de la CIM dans sa langue (français, allemand ou italien); la version anglaise étant la seule disponible au moment de l'étude. De plus, l'index alphabétique, véritable outil de recherche du code approprié, n'existait pas encore. D'autres difficultés sont inhérentes à la population des COMAI, comme la pluridisciplinarité des cas, qui rend plus compliqué la hiérarchisation des diagnostics. Ce problème est mis en lumière par le Tableau 4: la concordance est meilleure en considérant ensemble les 3 diagnostic principaux. A ce sujet,

une étude belge a trouvé un taux de discordance sur le choix du diagnostic principal de 31%⁸. Une autre source de différence d'appréciation commune peut être l'extraction du diagnostic qui entraîne parfois une variabilité de 40%¹². Les discordances de codages dépendent aussi de la qualité des informations consignées dans le dossier médical⁸.

Le fait que cette étude compare un codage fait par deux centres – donc deux formations –, à deux moments et deux endroits différents rend impossible de savoir lequel ou lesquels de ces facteurs jouent un rôle dans la discordance. Il s'agit certes d'une faiblesse des données dont on dispose. Par contre, cela permet à l'étude d'être le reflet strict de la réalité.

Les résultats par COMAI montrent qu'il est possible d'obtenir une concordance meilleure à tous les niveaux de précision.

On voit donc que la concordance de codage pourrait être améliorée notamment en formant les codeurs, en proposant des outils d'aide au codage (par exemple, un logiciel qui guiderait le codage ou, à défaut, l'index alphabétique), en adoptant une définition précise pour ce qu'est le premier diagnostic principal et en trouvant un consensus sur les discordances systématiques.

Niveau de concordance idéal

Quel niveau de précision doit être considéré comme l'idéal à atteindre? Certains auteurs préconisent l'utilisation des codes entiers de la CIM-10, afin d'obtenir des données cohérentes et riches de sens¹³. Cette proposition n'est pas acceptable ici car à une précision du code élevée est associé un taux de discordance également élevé ayant des conséquences sur l'occurrence des pathologies rencontrées. A quoi serviraient des données précises mais peu fiables? A l'inverse d'autres auteurs estiment qu'une fiabilité de deux tiers à une lettre + 1

chiffre est satisfaisante¹⁰. Il est vrai qu'une telle précision apporterait déjà une amélioration pour la description des causes de l'invalidité par rapport aux codes AI – encore en vigueur – pour lesquels, malgré 30 ans d'utilisation, aucun test de fiabilité n'a été publié. Les 66% suggérés sont atteints dans notre étude au niveau des groupes. Mais, les cas rencontrés aux COMAI ayant des pathologies recouvrant des domaines spécifiques, il est à craindre qu'une telle précision ne fournisse qu'une description médiocre. Il serait souhaitable de viser le niveau catégoriel, à condition d'améliorer la concordance y relative.

Formation des codeurs

La concordance entre juges de formations différentes, de même que la reproductibilité intra-juge, montrent des résultats globalement meilleurs pour le médecin possédant une pratique du codage. Cela ne signifie pas qu'une formation médicale soit une condition pour le codeur mais plutôt que les centres devraient disposer d'un collaborateur spécialement formé au codage, ce qui n'est pas le cas aujourd'hui puisque chaque médecin attribue les codes pour les patients qu'il examine. Une reproductibilité intra-juge de $\frac{3}{4}$ au niveau des groupes et de 50% au niveau des sous-catégories, pour les juges 2 et 3, est en partie le reflet de l'inexpérience. Une introduction au codage plus poussée – elle n'a été que d'une heure dans cette étude – aurait des chances d'améliorer ces résultats. Ici aussi, la mise à disposition des outils d'aide au codage dans la langue du personnel attribuant les codes aurait de chances d'améliorer la reproductibilité.

Limites de l'étude

Dans le cadre de l'assurance-invalidité, l'utilisation de la CIM-10 seule serait insuffisante pour les fins

d'évaluation, de gestion et de statistique. Elle pourrait être complétée par la Classification Internationale des Handicaps³ afin de disposer d'une description non seulement des causes de l'invalidité mais également de ses conséquences. Il conviendrait alors de faire une étude similaire pour la fiabilité du codage avec la Classification des handicaps aussi. Compte tenu des ressources limitées à la disposition de l'AI, il est raisonnable de penser que le remplacement progressif de ses propres codes par la CIM-10 soit envisageable dès à présent, d'autant plus que ses propres codes n'ont jamais été évalués quant à leur fiabilité et sont de peu d'intérêt pour l'épidémiologie ou la comptabilité de l'invalidité en Suisse.

Conclusion

Il est important de ne pas sous-estimer la complexité de cette tâche qu'est le codage et se rappeler qu'un codage incorrect produit des statistiques trompeuses, avec toutes les conséquences y relatives. Dans cette optique, il est sage de se contenter, pour l'analyse, d'un niveau de précision de codage s'arrêtant au groupe dans un premier temps. A ce niveau, la concordance trouvée entre le codage du CECO et celui des COMAI atteint un seuil jugé satisfaisant. Par la suite, l'introduction d'une série de mesures, visant à harmoniser les pratiques en matière de codage, permettrait d'atteindre le niveau de la catégorie. De plus, confier le codage à des personnes formées permettrait une spécialisation de la tâche et garantirait également une précision plus élevée. Cette même étude pourrait alors être répétée, dans ses grandes lignes, pour évaluer les bénéfices de telles interventions. L'Assurance-invalidité devrait envisager l'introduction généralisée de la CIM-10 pour le codage des diagnostics, causes d'invalidité.

Zusammenfassung**Zur Anwendung der internationalen Klassifikation der Krankheiten in der Invalidenversicherung: Eine Reliabilitätsstudie**

Der von der Invalidenversicherung (IV) zur Kodierung der Behinderungen verwendete hauseigene Kodierschlüssel hat sich als ungeeignet erwiesen, die verschiedenen Behinderungen zu erfassen. Als möglicher Ersatz wurde versuchsweise an einer Stichprobe von Fällen die internationale Klassifizierung der Krankheiten (ICD-10) angewendet. Der Zweck dieses Versuches war einerseits die Zuverlässigkeit dieser Kodierung zu testen, wobei die Behinderungen durch eine Medizinische Abklärungsstelle der IV (MEDAS) und ein zweites Mal durch das Evaluationszentrum kodiert wurden, sowie zu prüfen, ob dieser Kodierschlüssel auch durch Personen angewendet werden kann, die keine Ärzte sind. Entsprechend der Präzision der Diagnosen erfolgte die Analyse auf verschiedenen Ebenen: Kapitel, Gruppen, Kategorien und Subkategorien. Die Übereinstimmung der Kodierung der beiden Zentren schwankte zwischen 78% auf Stufe der Kapitel und 29% auf der Stufe mit der höchsten Präzision. Berücksichtigte man für diesen Vergleich jeweils alle drei Hauptdiagnosen zusammen, so erreichte die Übereinstimmung auf Stufe der Kapitel gar 91%. Die Reproduzierbarkeit (Kodierung und wiederholte Kodierung) war beim Arzt besser als bei der Ökonomin sowie der medizinischen Sekretärin. Die Übereinstimmung könnte durch die Schulung der Kodierer, durch das Anbieten von Kodierhilfen, durch eine eindeutige Definition der ersten Hauptdiagnose und durch eine gemeinsame Besprechung von systematisch auftretenden Unterschieden verbessert werden.

Summary**Use of the International Classification of Diseases for the disability insurance in Switzerland: A reliability study**

The International Classification of Diseases (ICD-10) was introduced in a pilot phase for coding of diagnoses in a sample of cases encountered in the Swiss Disability Insurance (SDI), as a substitute for the SDI codes, judged unsuitable for the description of disability. The goal of this study is to evaluate the reliability of use of these codes, by studying the coding reproductibility between the Medical Observation Centers (MOC) and the Evaluation and Coordination Center. A second goal was to evaluate the feasibility of coding by non-physicians. Reliability of diagnoses was examined by level of precision: chapters, groups, categories and sub-categories. The agreement between the two coding Centers varies from 78%, at the chapter level, to 29% at the most detailed level. This agreement is improved if taking into account the three main diagnoses for the comparison, reaching 91%. The results of the physician coder, concerning the precision and reliability of coding, are significantly superior to those of an economist coder and a medical secretary coder. To improve coding agreement, it is proposed to train coding professionals, provide coding tools, clarify which diagnosis is the main one, and seek a consensus for cases of systematic disagreement. The ideal level of precision should not be so high as to produce diagnostic codes with low reliability. It should, however, be precise enough to give a satisfactory description of the problems encountered.

Références

- 1 Wood VA, Wade DT, Hewer RL, Campbell MJ. The development of a disease classification system, based on the International Classification of Diseases, for use by neurologists. *Journal of Neurology, Neurosurgery, and Psychiatry* 1989; 52:449–458.
- 2 Hsia DC, Krushat WM, Fagan AB, Tebbut JA, Kusserow RP. Accuracy of diagnostic coding for medicare patients under the prospective-payment system. *N Engl J Med* 1988; 318:352–355.
- 3 Vez J-P, Vader JP, Paccaud F. Cross-coding from the Swiss Disability Insurance listing to the ICIDH. *Disability and Rehabilitation* 1995; 17:369–376.
- 4 Brämer GR. International statistical classification of diseases and related health problems tenth revision. *Rapp trimest statist sanit mond* 1988; 41:32–36.
- 5 Organisation Mondiale de la Santé. Classification statistique internationale des maladies et des problèmes de santé connexes (CIM-10) (10ème révision). Genève: Organisation Mondiale de la Santé, 1993.
- 6 Feinstein AR. Statistical indexes of association. In: Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: W.B. Saunders Company, 1985: 184–186.
- 7 Smith MW. Hospital discharge diagnoses: how accurate are they and their international classification of diseases (ICD) codes? *NZ Med J* 1989; 102:507–508.
- 8 Minvielle E, de Pourville G, Jeunemaitre A. P.M.S.I. Contrôle de la qualité du codage. *Gestions hospitalières* 1991; 302:17–22.
- 9 Okasha A, Seif El Dawla A. Reliability of ICD-10 research criteria: an arab perspective. *Acta Psychiatr Scand* 1992; 86:484–488.
- 10 Steinhilber H-Ch, Erdin A. The inter-rater reliability of child and adolescent psychiatric disorders in the ICD-10. *Journal of Child Psy-*

chology and Psychiatry; 32: 921–928.

- 11 *Crombie DL, Cross KW, Fleming DM.* The problem of diagnostic variability in general practice. *J Epidemiol Commun Health* 1992; 46:447–454.
- 12 *Colin C, Ecochard R, Messy P, Teboul F, Delahaye F, Landrison G, Morgon E.* Qualité de l'information médicale aux Hospices civils de Lyon: méthodes et résultats. *Santé Publique* 1992; 3:15–20.
- 13 *Hodgkin P, Metcalfe D.* Clustering diagnoses: a method of interpreting morbidity data. *Family Practice* 1984; 1:228–242.

Adresse pour correspondance

Dr. John-Paul Vader, MD, MPH
Institut de médecine sociale
et préventive
Université de Lausanne
rue du Bugnon 17
CH-1005 Lausanne
Fax +41 21 314 73 73
John-Paul.Vader@inst.hospvd.ch