

## Tests de validation des symptômes : vers plus d'objectivité et d'efficacité ?

Ces dernières années, la perception induite de prestations a fait l'objet de débats publics parfois vifs. Les méthodes d'instruction appliquées aux atteintes à la santé difficilement objectivables nécessiteraient des améliorations. Une étude a cherché à établir si les tests de validation des symptômes conviennent à la pratique de l'instruction et s'ils sont déjà utilisés. Elle montre que de nombreux tests de ce type sont décrits dans la littérature spécialisée, mais que seuls quelques-uns sont employés en pratique. Des réserves quant à un usage plus large ont cependant été formulées : certains estiment qu'ils ne sont pas assez précis et qu'ils posent des problèmes de définition dans l'examen de la crédibilité des troubles. Les tests de validation des symptômes ne devraient donc constituer qu'un des nombreux facteurs conditionnant l'exhaustivité de l'instruction.

Peter Rüesch

André Meichtry

René Schaffert

Jan Kool

ZHAW (Haute école zurichoise de sciences appliquées), département Santé

### Un nouvel instrument : les tests de validation des symptômes

Selon une étude récente (Ott, Bader et Wapf, 2007), on estime que 8 à 10% de rentes AI n'auraient pas été octroyées à bon escient. Les auteurs parlent à leur propos de « prestations non conformes aux objectifs de l'assurance ». Dans une petite partie de ces cas, on peut supposer qu'elles ont été octroyées parce que les assurés avaient fait de fausses déclarations. Les prestations AI non conformes aux objectifs de l'assurance sont allouées le plus souvent, comme on peut s'y attendre, pour des atteintes à la santé difficilement objectivables, telles que douleurs chroniques du dos sans cause somatique décelable, autres troubles douloureux, coup du lapin et dépressions. Ces tableaux cliniques ont fréquemment pour particularité de laisser une grande marge d'appréciation lorsqu'il s'agit

de déterminer l'incapacité de gain et le taux d'invalidité. Il serait donc intéressant de développer des instruments supplémentaires pour améliorer l'instruction par les centres d'observation médicale (COMAI) et les services médicaux régionaux (SMR), notamment en développant des normes applicables aux expertises qui portent sur des atteintes à la santé difficilement objectivables.

Les « tests de validation des symptômes » (TVS) jouent là un rôle important. Il s'agit de procédures visant à vérifier la plausibilité des troubles décrits par les assurés dans le cadre de l'instruction. Il en existe deux types :

- les tests standardisés, basés généralement sur un questionnaire rempli par les assurés eux-mêmes ou par les experts ;
- les directives, qui comprennent plusieurs critères permettant de contrôler la crédibilité des troubles décrits par les expertisés.

Les TVS ont un aspect « magique », un peu comme un détecteur de mensonges ; il peut sembler tentant de disposer d'un instrument qui promet de vérifier, sous une forme simple et objective, la plausibilité des troubles allégués. Mais ces tests sont-ils vraiment fiables et utilisables dans le cadre de l'instruction ? Et où en sont les connaissances scientifiques dans ce domaine ? Une étude réalisée à la demande de l'Office fédéral des assurances sociales (Kool, Meichtry, Schaffert, Rüesch, 2008) a cherché à y répondre. Ses auteurs ont, d'une part, procédé à une recherche bibliographique des TVS dans la littérature scientifique et, d'autre part, réalisé des entretiens et une enquête auprès des experts exerçant dans les COMAI et des collaborateurs des SMR. Nous présentons ici les principaux aspects théoriques du développement et de l'utilisation des TVS qui peuvent être importants pour la pratique de l'instruction.

### Critères de qualité des tests

La littérature scientifique sur les tests de validation des symptômes est abondante. Une revue systématique de la littérature publiée de 1997 à 2007 a donné 1100 références : 570 pour les cinq dernières années (2003-2007), dont 340 depuis 2005. Le sujet est connu depuis longtemps dans l'espace anglophone. Les tests en anglais ne pouvant pas être transposés tels quels aux langues nationales suisses, la recherche effectuée dans le cadre de l'étude s'est limitée à ceux dont il existe une version allemande ou française. Par ailleurs, les TVS sont employés dans des environnements très divers (expertises en lien avec la capacité de travail, examens de médecine légale...); la

recherche a donc été limitée aux thèmes en relation avec l'AI. Au total, les auteurs ont trouvé des références sur 30 TVS en allemand. La recherche n'a pas donné de résultats pour l'espace francophone.

Etant donné le nombre de tests à disposition, des critères concrets sont particulièrement indiqués pour juger de leur qualité. En ce qui concerne leur valeur pratique, ces critères sont les mêmes pour les TVS que pour d'autres tests, par exemple psychologiques: la validité, la fiabilité et l'objectivité. De ce point de vue, les TVS existants présentent un certain nombre de points critiques qui doivent être pris en compte pour leur utilisation et leur développement. Commençons par une brève définition des trois critères:

- **validité**: un test doit mesurer ce qu'il est censé mesurer. Par exemple, un test de mathématiques qui, pour la résolution des problèmes, présupposerait de très bonnes compétences en lecture n'aurait qu'une validité limitée;
- **fiabilité**: un test doit être reproductible et différencié en fonction des sujets qui le passent, de façon à fournir des résultats fiables. La précision des mesures est étroitement liée à la fiabilité;
- **objectivité**<sup>1</sup>: ce critère, qui s'applique à l'utilisation concrète du test, signifie que celui-ci devrait toujours fournir les résultats les plus constants possible, indépendamment de la personne qui le fait passer et des conditions de son passage.

A noter que les critères brièvement définis ici découlent les uns des autres: sans objectivité dans l'application, la fiabilité n'est pas possible;

sans fiabilité, il ne peut pas non plus y avoir de validité.

### Validité: que mesurent les tests de validation des symptômes?

Les TVS doivent fournir une indication sur la plausibilité des troubles décrits par les assurés. A ce propos, les notions d'*exagération* et de *simulation* sont particulièrement importantes. Dans la littérature spécialisée, la notion d'exagération est définie comme l'accentuation ou l'extension des troubles: le patient renforce les symptômes réellement présents afin d'atteindre un objectif (rente, mesure, etc.). La simulation est l'imitation délibérée et réfléchie des symptômes, dans un but précis, ou la description mensongère des troubles. Les deux comportements se manifestent souvent par une incohérence entre les capacités ou les performances qui sont observées (pendant l'examen) et celles qui sont attendues (sur la base des troubles décrits). Les TVS sont donc censés mesurer l'exagération ou la simulation dans un domaine précis (par exemple les troubles neurologiques).

Pour ce qui est de la validité, il existe cependant un problème majeur: les incohérences entre performances observées et performances attendues peuvent exprimer non seulement des phénomènes d'exagération ou de simulation, mais aussi des atteintes à la santé réelles. Dans la littérature spécialisée, on compte

parmi ces troubles certaines maladies psychiques, à savoir les troubles somatoformes et le trouble factice. Comment donc faire la différence entre exagération ou simulation d'une part et troubles ou maladie d'autre part? Deux critères peuvent s'avérer utiles (cf. tableau **T1**): le fait que la personne est motivée par un intérêt extérieur (obtenir une rente, voir une sanction prononcée, etc.) et la conscience qu'elle a de son comportement.

On postule que plus le comportement est motivé par des intérêts extérieurs et plus il est conscient, plus grande est la probabilité d'avoir affaire à une exagération ou à une simulation. Les deux critères – motivation et conscience – ne peuvent cependant ni être mesurés instrumentalement ni objectivés. L'expert dispose en fait d'une certaine marge d'appréciation pour déterminer la motivation et le degré de conscience de la personne qu'il examine. Etant donné ces problèmes théoriques, il faut, pour déterminer s'il s'agit d'une exagération ou d'une simulation, un diagnostic différentiel fondé, capable d'exclure avec une grande vraisemblance d'autres explications du comportement de l'assuré.

### Peut-on se fier aux mesures des tests de validation des symptômes?

En pratique, on attend d'un TVS qu'il mette en évidence, de la manière la plus fiable possible, les

### Différences entre simulation / exagération et trouble factice / troubles somatoformes / troubles dissociatifs

T1

Diagnostic	Conscience du mode d'élaboration des symptômes	Motivation du comportement
<b>Simulation/exagération</b>	Délibéré, dirigé (« conscient »)	Réfléchi, clair (« conscient »)
<b>Trouble factice</b>	Délibéré, dirigé (« conscient »)	Non réfléchi, imprécis (« inconscient »)
<b>Troubles somatoformes et troubles dissociatifs</b>	Involontaire, non dirigé (« inconscient »)	Non réfléchi (« inconscient »)

<sup>1</sup> Le critère d'objectivité étant très exigeant, il est difficile d'y répondre du point de vue scientifique au niveau théorique. Dans le présent contexte, une construction de test indépendante de la mesure semble illusoire a priori; de ce fait, la notion plus modeste d'intersubjectivité doit être préférée à celle d'objectivité.

troubles qui sont exagérés ou simulés et qu'il les distingue des troubles réels. Il doit donc donner des résultats positifs dans le premier cas et des résultats négatifs dans le second. Mais il est très difficile de répondre à ces deux exigences à la fois dans la construction d'un test, comme le montre l'exemple tiré de leur utilisation dans un autre contexte, le dépistage médical (cf. tableau T2). Le pourcentage de malades détectés comme tels par le test, c'est-à-dire pour lesquels celui-ci a donné un résultat positif, est la «sensibilité» (c'est-à-dire:  $a/(a+c)$ ), tandis que le pourcentage des personnes en bonne santé qui obtiennent un résultat négatif est appelé «spécificité» ( $d/(b+d)$ ).

La difficulté est que sensibilité et spécificité sont liées: la sensibilité d'un test est d'autant plus grande que sa spécificité est faible. En pratique, les développeurs doivent donc trouver entre ces deux grandeurs le compromis optimal pour chaque usage. Dans le contexte médical, ce qui est le cas par exemple pour le diagnostic précoce d'une maladie grave, la priorité est d'optimiser la sensibilité, même si le test donne des faux positifs chez un certain nombre de personnes. Dans le contexte de la validation des symptômes, en revanche, le plus important est la spécificité car, étant donné les conséquences pour la personne concernée (le risque étant de l'accuser à tort de simulation), il vaut mieux éviter les faux positifs.

Le développement de TVS fiables bute donc sur deux écueils. Tout d'abord, il n'existe pas d'«étalon-or» dans le domaine de la validation des symptômes: aucun instrument ne permet de faire, de manière fiable, la distinction entre vérité et mise en scène (exagération ou simulation) de troubles. En effet, presque tous les tests décrits dans la littérature spécialisée n'ont pas été mis au point dans un environnement réel mais en laboratoire, avec des personnes qui en quelque sorte «simulent la simu-

lation». Pour nombre d'entre eux, rien ne prouve qu'il est possible de transposer ces résultats de laboratoire à la pratique d'un service d'observation régional.

Ensuite, la sensibilité et la spécificité d'un TVS dépendent aussi de la fréquence (prévalence) des phénomènes d'exagération et de simulation dans l'ensemble de la population des assurés examinés dans un centre d'observation. Si cette fréquence est très faible, la probabilité qu'une personne ayant obtenu un résultat positif au test présente effectivement le phénomène recherché – simulation par exemple – est également très faible. Cette probabilité constitue la «valeur prédictive positive» (VPP) du test. Celle-ci est illustrée par un exemple chiffré (cf. tableau T3) pour un test présentant une sensibilité parfaite (1,0) et une spécificité quasi parfaite (0,95).

Si la prévalence de la simulation dans la population d'assurés est de 15%, la VPP est de 78%. Mais si elle n'est que de 5%, la VPP est de 52%, avec par conséquent un grand nombre de faux diagnostics. Autrement

dit: si la fréquence des phénomènes d'exagération et de simulation dans la population est faible, même un test très précis risque de ne fournir que des indications relativement imprécises quant à la probabilité qu'un assuré exagère ou simule réellement. La prévalence de ces phénomènes chez les assurés AI n'est pas connue, mais, d'après l'étude d'Ott, Bade et Wapf citée, les chiffres sont plutôt bas, c'est-à-dire nettement inférieurs à 10%.

### Comment les tests de validation des symptômes sont-ils utilisés ?

L'intersubjectivité des TVS peut, de manière générale, être considérée comme un fait: lors de leur développement, on vérifie par exemple si des utilisateurs différents obtiennent les mêmes résultats. Mais ce que l'on sait moins, c'est comment ces tests sont utilisés en pratique, notamment en Suisse. L'étude réalisée par l'OFAS comprenait des entretiens personnels et une enquête par ques-

## Sensibilité et spécificité d'un test

T2

Test	Maladie (p.ex. cancer du sein)	
	présent (+)	absent (-)
positif (+)	vrai positif (a)	faux positif (b)
négatif (-)	faux négatif (c)	vrai négatif (d)

## Dépendance entre fiabilité d'un test et fréquence du phénomène à étudier

T3

Test	Phénomène : prévalence = 15%			Phénomène : prévalence = 5%		
	+	-	Total	+	-	Total
+	150	42	192	50	47	97
-	0	808	808	0	903	903
<b>Total</b>	150	850	1000	50	950	1000
Sensibilité = 100,00%			Sensibilité = 100,00%			
Spécificité = 95,06%			Spécificité = 95,05%			
Valeur prédictive positive (VPP) = 78,13%			Valeur prédictive positive (VPP) = 51,55%			

tionnaire auprès des experts exerçant dans les COMAI et des collaborateurs des SMR. La majorité d'entre eux jugent faible la fréquence (prévalence) des phénomènes d'exagération et de simulation dans le cadre de l'instruction, la plupart estimant que la simulation ne constitue pas un problème majeur dans leur travail. Ils expliquent les taux élevés rapportés dans les études scientifiques par la spécificité des populations étudiées et par la non-prise en compte des facteurs sociaux dans la genèse des troubles chroniques. Il paraît difficile de conclure avec certitude à un phénomène de simulation, puisque la moitié des sondés estime qu'il est rarement possible de poser un diagnostic certain dans ce domaine (graphique **G1**). Il serait plus facile de repérer les assurés qui exagèrent leurs troubles.

Le questionnaire a montré par ailleurs que peu de TVS sont employés à l'heure actuelle dans le cadre de l'instruction. De plus, la façon

de les utiliser n'est pas uniforme, bien que l'on trouve des directives très complètes dans la littérature spécialisée. Celles-ci définissent plusieurs critères qui doivent être satisfaits pour que des soupçons d'exagération ou de simulation soient confirmés. Parmi les directives reconnues, citons celles de Bianchini et al. (2005) pour les douleurs chroniques. Les auteurs proposent les cinq critères (A-E) ci-dessous, dont les deux plus spécifiques (A et E) doivent toujours être satisfaits pour que l'on puisse parler de simulation avec une certitude suffisante. Les autres critères (B, C et D) servent davantage à déterminer le degré de simulation :

- A: preuve de l'existence de facteurs extérieurs incitant à simuler (instruction en vue d'une rente, par exemple);
- B: indices d'une exagération du handicap durant l'examen clinique (tels que performances inférieures aux résultats attendus ou incohérences);

- C: indices d'une exagération du handicap dans les tests cognitifs ou neuropsychologiques;
- D: indices d'une exagération du handicap dans la description des symptômes et des troubles;
- E: impossibilité d'expliquer totalement les résultats des critères A à D par des facteurs psychiatriques, neurologiques ou développementaux.

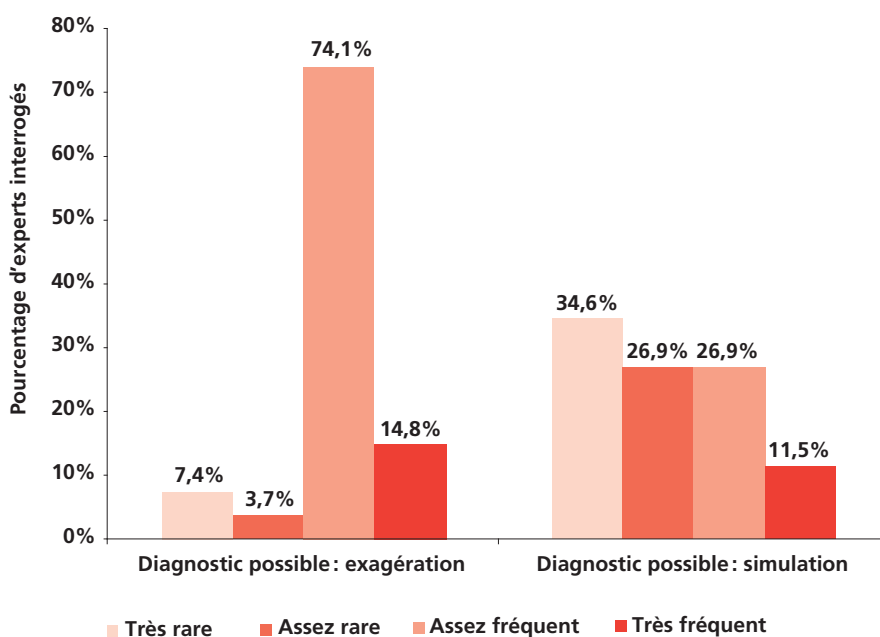
## Conclusion

Dans le cadre d'une instruction de l'AI, les tests de validation des symptômes peuvent constituer un élément important de l'examen systématique visant à établir la réalité des troubles décrits par les assurés. Mais ils ne doivent en aucun cas remplacer des examens complets. La liste des TVS figurant dans la littérature scientifique est relativement longue, du moins pour l'espace germanophone. On peut toutefois exprimer des réserves quant à un usage non critique dans les conditions concrètes d'un centre d'observation (COMAI par exemple), comme nous avons tenté de le montrer dans le présent article. La majorité de ces tests n'ont pas encore suffisamment fait leurs preuves sur des échantillons d'assurés. En outre, si la prévalence des phénomènes d'exagération et de simulation est faible dans la population des bénéficiaires de rente AI, ils risquent de générer un trop grand nombre d'erreurs d'appréciation, ce qui serait inacceptable du point de vue éthique. Enfin, la base théorique sur laquelle ils se fondent n'est pas encore convaincante, car les phénomènes en question sont difficiles à distinguer de certains troubles pathologiques véritables.

Les TVS ne sont cependant pas sans intérêt. Ils peuvent contribuer à harmoniser la procédure de validation des symptômes dans la pratique de l'instruction. Ils obligent leurs utilisateurs à rechercher des critères mesurables pour le diagnostic d'exa-

**Fiabilité du diagnostic d'exagération et de simulation : point de vue des experts des COMAI et des SMR (N=27, d'après Kool et al., 2008)**

**G1**



gération ou de simulation. La littérature spécialisée fournit en outre toute une série de directives utiles pour vérifier la plausibilité de différents tableaux cliniques, dont nous recommandons l'application à la pratique des COMAI et des SMR. Cette démarche serait en fin de compte dans l'intérêt des assurés, qui pourraient ainsi compter sur une procédure d'instruction transparente et uniforme. Par ailleurs, malgré la fiabilité encore douteuse des TVS, leur non-utilisation ne rendra pas le diagnostic plus sûr; cette option n'est donc pas plus responsable. A la différence des tests – dont on peut contrôler la validité diagnostique (sensibilité, spécificité) –, la fiabilité diagnostique des autres méthodes est inconnue. Nous recommandons par conséquent d'utiliser davantage les tests de validation des symptômes et de soumettre à un examen scientifique quelques procédures choisies auprès d'échantillons d'assurés AI réels.

## Bibliographie

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: lessons from cognitive malingering research. *Spine Journal*, 5(4), 404-417.

Kool, J.; Meichtry, A.; Schaffert, R.; Rüesch, P. (2008). Validation des symptômes dans l'instruction AI. Berne. Office fédéral des assurances sociales.

Ott, W., Bade, S., & Wapf, B. (2007). Prestations non conformes aux objectifs dans l'AI. Berne. Office fédéral des assurances sociales.

Peter Rüesch, docteur en psychologie FSP, responsable du centre de sciences de la santé à la Haute école zurichoise de sciences appliquées (Hochschule für Angewandte Wissenschaften, ZHAW), département Santé.

Mél. : rech@zhaw.ch

André Meichtry, MPTSc, physiothérapeute, professeur de physiothérapie à la ZHAW, département Santé.

Mél. : mcdm@zhaw.ch

René Schaffert, lic. phil., sociologue, collaborateur scientifique du service Sciences de la santé de la ZHAW.

Mél. : shar@zhaw.ch

Jan Kool, PhD, physiothérapeute, responsable Recherche et développement en physiothérapie à la ZHAW, département Santé.

Mél. : kool@zhaw.ch