

Les comparaisons internationales d'état de santé subjectif sont-elles pertinentes ? Une évaluation par la méthode des vignettes-étalons

Salim Lardjane* et Paul Dourgnon**

Lorsque les modalités de réponse à une question d'évaluation subjective sont utilisées différemment par des répondants, on dit que celles-ci sont affectées par un « biais d'item » ou par un effet de type *DIF* (*Differential Item Functioning*). Les réponses doivent alors être corrigées avant de pouvoir être utilisées pour formuler des hypothèses comparatives sur l'état de santé subjectif de sous-populations.

Nous présentons dans ce travail une méthode de détection et de correction d'effets *DIF* dans les auto-évaluations de santé subjective. Cette méthode est basée sur l'évaluation par les individus d'états de santé correspondant à des situations fictives, appelées vignettes-étalons.

Elle consiste à « étalonner » la réponse d'un individu à une question subjective par les évaluations qu'il fait de situations décrites dans différentes vignettes. Elle suppose que tous les individus se réfèrent à une même dimension de santé en répondant aux questions subjectives, qu'ils évaluent de façon comparable les vignettes et leur propre situation, que les vignettes puissent être ordonnées *a priori* et sans ambiguïté par l'analyste, selon la dimension de santé étudiée.

Un biais *DIF* potentiel peut être identifié en comparant la répartition des évaluations des vignettes dans deux sous-groupes d'individus. Il peut être corrigé en positionnant la réponse de chaque individu à la question d'évaluation subjective par rapport à ses évaluations des situations décrites dans les vignettes. La prise en compte de problèmes de cohérence conduit à construire des intervalles d'incertitude pour les valeurs prises par la variable ainsi obtenue.

La méthode proposée est illustrée par la mise en évidence d'un possible effet *DIF* dans l'auto-évaluation de la douleur physique, selon l'origine géographique des échantillons issus de l'enquête *Share* 2004, qui nous conduirait en particulier à modifier notre comparaison des états de santé subjectifs des échantillons suédois et néerlandais.

* Crest, Laboratoire de Statistique d'Enquêtes et GSF-National Research Center for Environment and Health, Institute of Biomathematics and Biometry, Neuherberg, Allemagne

** Ides, Institut de Recherche et Documentation en Économie de la Santé

La mesure de la santé des individus est essentielle pour évaluer les politiques publiques. Toutefois, il n'est pas toujours possible de mesurer les différentes dimensions de la santé étudiées à l'aide de procédés de mesures physiques ou d'avis de médecins, en raison de coûts trop élevés ou d'impossibilités pratiques. On a alors recours à des mesures subjectives de santé, typiquement obtenues à l'aide de questionnaires d'auto-évaluation (Falissard, 2001). Ces derniers peuvent, par exemple, être construits, lors de la conception de l'étude, en sélectionnant une ou plusieurs dimensions de santé, telles la douleur physique, la mobilité, la capacité cognitive, dans la classification ICIDH-2 (*International Classification of Impairments, Disabilities and Handicaps*) mise au point par l'OMS (World Health Organization, 1980 et 1999), puis en associant à chacune d'entre elles une ou plusieurs questions d'auto-évaluation. En répondant à une question d'auto-évaluation, un individu évalue lui-même son état de santé, pour une dimension de santé donnée. Le questionnaire utilisé pour l'enquête *Share* (cf. encadré 1) illustre cette approche.

Dans le questionnaire *Share*, une question d'auto-évaluation est associée à chaque dimension de santé retenue et cinq modalités de réponse sont proposées pour chaque question d'auto-évaluation. Les dimensions de santé considérées et les questions d'auto-évaluation correspondantes sont directement reprises des questionnaires conçus par l'OMS pour l'Enquête mondiale sur la Santé 2002-2003. Dans ce travail, nous traiterons de la question d'auto-évaluation associée à la *douleur physique* (1).

Le problème de comparabilité des réponses dans les enquêtes de santé

Supposons que les répondants partagent les mêmes attentes ou une même norme de santé pour une dimension de santé donnée. Dans ce cas, on a de bonnes raisons d'avancer que deux individus ayant répondu par une même modalité à une question d'auto-évaluation associée ont des niveaux de santé subjective relativement proches. Dans une telle situation, un individu ayant répondu « léger » à la question d'auto-évaluation précédente sur la douleur physique (cf. *supra*) se perçoit en meilleure santé, de ce point de vue, qu'un individu ayant répondu « moyen » à cette même question.

Supposons à présent que deux répondants aient des attentes ou des normes très différentes pour

une même dimension de santé. Dans ce cas, les réponses obtenues ne permettent pas nécessairement de comparer l'état de santé subjectif des individus. Ainsi, une personne atteinte de troubles respiratoires peut utiliser la modalité « léger » en réponse à une question d'auto-évaluation sur sa mobilité pour signifier qu'elle ne pense pas parvenir à effectuer 1 km sans s'es-souffler, alors qu'un coureur de demi-fond peut utiliser cette même modalité « léger » pour signifier qu'il ne pense pas parvenir à effectuer ses 20 km de course hebdomadaires habituels. Dans une telle situation, certes extrême, les modalités de réponse sélectionnées ne permettent pas de comparer la mobilité subjective des deux individus.

Cet exemple illustre le fait que les modalités de réponse à la question d'auto-évaluation peuvent être utilisées différemment par chacun et donc qu'elles peuvent « fonctionner » différemment selon les répondants. De façon plus formelle, on dit alors que les modalités de réponse – les *items* en psychométrie – sont affectés par un « biais d'item » ou d'un effet de type « DIF » (*Differential Item Functioning*, cf. Nunnally et Bernstein, 1994 ; Falissard, 2001).

La présence d'un effet DIF affectant les réponses à une question d'auto-évaluation est toujours une possibilité théorique mais, *sans plus d'information, cette possibilité doit être écartée* pour des raisons tant éthiques (principe d'impartialité) que scientifiques (principe de parcimonie). Il est donc fondé, sans plus d'information, de supposer *qu'il n'y a pas* d'effet DIF et donc pas de problème de comparabilité des réponses obtenues à une question d'auto-évaluation.

Diverses études comparatives de morbidité et de santé (Murray et Chen, 1992 ; Salomon *et al.*, 2004 ; Sen, 2002, entre autres) ont toutefois montré que de tels problèmes de comparabilité des réponses étaient potentiellement présents dans les enquêtes sur la santé subjective, notamment lorsqu'on souhaitait comparer des échantillons issus de pays ou régions différents pour une ou plusieurs dimensions de santé, ce qui est la situation typique des enquêtes effectuées sous l'égide d'organisations internationales, telles l'ONU, l'OMS ou l'Union européenne. Ces travaux ont également montré que l'occultation des

1. Question d'auto-évaluation pour la dimension de santé « douleur physique » (questionnaire *Share* 2004) : « Dans l'ensemble, au cours des 30 derniers jours, quel niveau de douleurs physiques avez-vous ressenti ? »

Aucun Léger Moyen Grave Extrême

problèmes de comparabilité des réponses pouvait conduire à des *contresens graves* sur l'état de santé des populations, notamment des plus défavorisées économiquement. Les résultats de ces études, confrontés aux considérations précédentes sur l'effet *DIF*, plaident en faveur du recueil *d'informations supplémentaires* en complément des questions d'auto-évaluation lors d'enquêtes de santé subjective, afin de pouvoir détecter d'éventuels problèmes de comparabilité des réponses.

La méthode présentée dans ce travail est une illustration de cette approche. En exploitant des questions complémentaires associées à la question d'auto-évaluation sur la douleur physique dans le questionnaire *Share*, nous mettons en évidence, sous certaines hypothèses, un effet *DIF* et montrons comment le prendre en compte lors de comparaisons entre échantillons nationaux.

Les vignettes-étalons : un outil pour détecter et corriger d'éventuels problèmes de comparabilité

Ce travail se situe dans la continuité de méthodes récentes d'analyse et de traitement des effets *DIF* dans les enquêtes de santé, développées autour de l'OMS et du département de Sciences politiques de l'Université de Harvard depuis le début des années 2000. Ces méthodes reposent sur l'inclusion d'éléments supplémentaires, appelés *vignettes d'ancrage* (*Anchoring Vignettes, Anchors*), que nous proposons d'appeler également *vignettes-étalons*, dans les questionnaires de santé subjective. C'est l'intérêt récent pour ces méthodes et leur utilité potentielle dans une optique de comparaisons internationales qui explique que de telles vignettes aient été incluses dans divers questionnaires de santé de l'OMS et dans le questionnaire *Share*.

Le terme *vignette* est utilisé en sciences sociales pour désigner toute *description d'une situation fictive* soumise à un individu, à laquelle on lui demande de réagir ou qu'on lui demande d'évaluer. L'utilisation de celles-ci remonte aux années 1950, avec des travaux en anthropologie, psychologie et sociologie, mais également en droit et en sciences de l'éducation (Alexander et Becker, 1978 ; Salomon *et al.*, 2004). Des vignettes ont, par ailleurs, été récemment utilisées en médecine, notamment en lien avec la formation des médecins et des infirmiers (Salomon *et al.*, 2004).

On appellera spécifiquement *vignette d'ancrage* ou *vignette-étalon* associée à une question d'auto-évaluation toute vignette soumise aux individus enquêtés, portant sur la même dimension de santé que la question d'auto-évaluation, et qu'on demande aux individus d'évaluer sur l'échelle utilisée pour la question d'auto-évaluation, en réponse à une question aussi proche que possible de celle utilisée pour l'auto-évaluation (cf. encadré 2 pour les trois vignettes-étalons associées à la question d'auto-évaluation portant sur la douleur physique dans le questionnaire *Share*).

Dans l'optique d'une utilisation pour détecter et corriger un effet *DIF*, les vignettes-étalons et le protocole d'enquête doivent être conçus de façon à ce que les hypothèses suivantes puissent raisonnablement être considérées comme vérifiées.

H1. *Hypothèse d'équivalence des vignettes* : tous les répondants comprennent de la même façon la situation décrite dans chacune des vignettes.

H1 signifie que chaque vignette suffit à décrire de façon satisfaisante une situation qui est la même pour tous les individus enquêtés. Dans le cas de la douleur physique, cela revient à sup-

Encadré 1

QUELQUES ÉLÉMENTS SUR L'ENQUÊTE SHARE 2004 ET LES DONNÉES UTILISÉES

L'enquête *Share 2004 – Survey of Health, Ageing and Retirement in Europe* (Börsch-Supan et Jürges, 2005) – a été réalisée dans 11 pays européens, donc en différentes langues. Cette enquête a été explicitement conçue dans le but de répondre à des problématiques de comparaisons internationales. Les divers questionnaires nationaux ont été obtenus à partir d'un questionnaire générique en anglais, qui a été traduit question par question. Dans ce travail,

on limite le champ de l'étude aux individus enquêtés ayant répondu à la question d'auto-évaluation sur la douleur physique et aux questions de vignettes associées, lesquelles n'ont été posées que dans huit pays : Allemagne, Belgique, Espagne, France, Grèce, Italie, Pays-Bas, Suède, avec une distinction entre Belgique francophone et Belgique néerlandophone. On dispose donc finalement de neuf échantillons.

poser que les individus évaluent bien une même situation en terme de douleur lors de l'évaluation d'une vignette. On mesure l'importance d'une bonne conception des vignettes et d'une traduction de qualité pour que cette hypothèse soit approximativement vérifiée dans le cadre d'enquêtes plurilingues.

H2. *Hypothèse d'équivalence des modalités de réponse* : chaque répondant utilise les modalités de réponse de la même façon lors de l'auto-évaluation et lors de l'évaluation des différentes vignettes.

H2 autorise à *comparer* formellement, à partir des réponses obtenues, l'auto-évaluation par un individu donné de son état de santé à son évaluation des différentes vignettes, et conditionne donc pour une large part l'utilisation des vignettes-étalons pour la mise en évidence et la correction d'un effet *DIF*. Une bonne formulation de la question et des modalités de réponse est essentielle pour que cette hypothèse puisse être vérifiée. Celle-ci *n'implique pas* que les *différents* répondants utilisent les modalités de réponse de la même façon, auquel cas il n'y aurait pas d'effet *DIF*.

Divers auteurs augmentent l'hypothèse d'équivalence des vignettes d'une hypothèse por-

tant sur l'unidimensionnalité des évaluations ou la font implicitement – nous avons plutôt choisi de l'énoncer séparément en raison de l'usage qui en sera fait et de l'importance de l'hypothèse d'équivalence des vignettes, telle qu'énoncée ci-dessus, pour les comparaisons internationales.

H3. *Hypothèse d'unidimensionnalité* : l'auto-évaluation et l'évaluation des différentes vignettes se rapportent à une même dimension de santé pour tous les répondants.

Cette hypothèse permet d'une part, de considérer que les comparaisons formelles déduites des réponses d'un individu portent effectivement sur une même dimension de santé et d'autre part, de considérer qu'on a bien affaire à une même dimension de santé pour les différents répondants. Elle est *fondamentale*, et conditionne toute analyse effectuée à partir des auto-évaluations et, le cas échéant, des évaluations des vignettes associées.

Pour ce qui concerne la douleur physique dans l'enquête *Share* (cf. encadré 3), les hypothèses précédentes sont vérifiées dans la situation où chaque individu enquêté se « met à la place » de l'individu décrit dans chacune des vignettes et évalue la douleur physique correspondante,

Encadré 2

**VIGNETTES ASSOCIÉES À LA DIMENSION DE SANTÉ
« DOULEUR PHYSIQUE » - QUESTIONNAIRE SHARE 2004**

Première vignette (vignette V_1) associée à la dimension de santé *douleur physique* dans le cadre de l'enquête *Share* 2004.

V_1 : Paul a un mal de tête une fois par mois qui diminue après qu'il ait pris un cachet. Pendant qu'il a mal à la tête, il peut mener ses activités quotidiennes.

Question associée :

En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Paul a-t-il éprouvé ?

Aucun	Léger	Moyen	Grave	Extrême
<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Deuxième vignette (vignette V_2) associée à la dimension de santé *douleur physique* dans le cadre de l'enquête *Share* 2004.

V_2 : Henri a mal dans tout son bras droit et son poignet pendant sa journée de travail. Cela est partiellement atténué la soirée lorsqu'il ne travaille plus devant l'ordinateur.

Question associée :

En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Henri a-t-il éprouvé ?

Aucun	Léger	Moyen	Grave	Extrême
<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

Troisième vignette (vignette V_3) associée à la dimension de santé *douleur physique* dans le cadre de l'enquête *Share* 2004.

V_3 : Charles a mal aux genoux, aux coudes, aux poignets et aux doigts, et la douleur est presque continuellement présente. Bien que les médicaments aident, il ne se sent pas bien lorsqu'il se déplace, qu'il tient ou soulève quelque chose.

Question associée :

En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Charles a-t-il éprouvé ?

Aucun	Léger	Moyen	Grave	Extrême
<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄	<input type="checkbox"/> ₅

et uniquement la douleur physique correspondante, comme si c'était lui-même qui la subissait.

Les hypothèses précédentes suffisent à justifier l'utilisation des vignettes-étalons pour la *détection* d'un effet *DIF*. Lorsqu'il y a plus d'une

Encadré 3

**DÉFINITION D'UNE VERSION DES AUTO-ÉVALUATIONS
CORRIGÉE DE L'EFFET *DIF*
POUR LES RÉPONDANTS DONT LES RÉPONSES SONT COHÉRENTES
- DIMENSION DE SANTÉ « DOULEUR PHYSIQUE » - *SHARE* 2004**

La réponse à la question d'auto-évaluation possède cinq modalités sur une échelle ordinale. À chaque individu i , on soumet trois vignettes à évaluer sur la même échelle (cf. encadré 2).

L'auto-évaluation de l'individu i est notée $Y(i)$. L'évaluation de la vignette j par l'individu i est notée $Z(i,j)$. Les vignettes associées à la question d'auto-évaluation sont numérotées de V_1 à V_3 par niveau de gravité strictement croissant, selon l'ordre privilégié $V_1 < V_2 < V_3$.

Avec cette convention, les réponses de l'individu i sont cohérentes si et seulement si $Z(i,1) < Z(i,2) < Z(i,3)$.

Si c'est le cas, on pose

$$C(i) = 1 \text{ si } Y(i) < Z(i,1)$$

$$C(i) = 2 \text{ si } Y(i) = Z(i,1)$$

$$C(i) = 3 \text{ si } Z(i,1) < Y(i) < Z(i,2)$$

$$C(i) = 4 \text{ si } Y(i) = Z(i,2)$$

$$C(i) = 5 \text{ si } Z(i,2) < Y(i) < Z(i,3)$$

$$C(i) = 6 \text{ si } Y(i) = Z(i,3)$$

$$C(i) = 7 \text{ si } Y(i) > Z(i,3)$$

C est donc une variable ordinale à 7 modalités. Plus généralement, si on avait eu J vignettes, on aurait pu construire de la même façon une variable ordinale ayant $2J + 1$ modalités.

Si $C(i)$ est *pair*, cela signifie que la *même* modalité de réponse a été sélectionnée pour la question d'auto-

évaluation et l'une des vignettes associées ; le rang de cette dernière dans la série de vignettes ordonnée par gravité strictement croissante est $C(i)/2$. Si $C(i)$ est *impair*, cela signifie que la modalité sélectionnée en réponse à la question d'auto-évaluation n'a été retenue pour aucune des vignettes ; l'état de santé subjectif de l'individu se positionne, en termes de rang dans la série des vignettes ordonnée par gravité strictement croissante, entre $(C(i) - 1)/2$ et $(C(i) + 1)/2$.

À titre d'illustration, comparons l'état de santé subjectif de deux individus dont les réponses sont cohérentes, ayant tous deux déclaré un niveau de douleur physique « moyen ».

Soit i un répondant pour lequel $Y(i)$ = « moyen » et pour lequel $Z(i,1)$ = « léger », $Z(i,2)$ = « grave » et $Z(i,3)$ = « extrême ». Comme $Z(i,1) < Z(i,2) < Z(i,3)$, les réponses de l'individu i sont cohérentes. D'autre part, comme $Z(i,1) < Y(i) < Z(i,2)$, on a $C(i) = 3$.

Soit par ailleurs m un répondant pour lequel $Y(m)$ = « moyen » et pour lequel $Z(m,1)$ = « léger », $Z(m,2)$ = « moyen » et $Z(m,3)$ = « extrême ». Comme $Z(m,1) < Z(m,2) < Z(m,3)$, les réponses de l'individu m sont également cohérentes. D'autre part, comme $Y(m) = Z(m,2)$, on a $C(m) = 4$.

On en déduit que le niveau de douleur physique ressenti par m est plus élevé que le niveau de douleur physique ressenti par i , bien qu'ils aient tous deux répondu de la même façon à la question d'auto-évaluation sur la douleur physique.

Tableau A
Deux cas de cohérence

	Aucun	Léger	Moyen	Grave	Extrême
Premier cas					
Réponse à la question de l'enquête <i>Share</i>			$Y(i)$		
Réponse à la première vignette V_1		$Z(i,1)$			
Réponse à la deuxième vignette V_2				$Z(i,2)$	
Réponse à la troisième vignette V_3					$Z(i,3)$
Valeur de $C(i)$	$C(i) = 3$				
Deuxième cas					
Réponse à la question de l'enquête <i>Share</i>			$Y(m)$		
Réponse à la première vignette V_1		$Z(m,1)$			
Réponse à la deuxième vignette V_2			$Z(m,2)$		
Réponse à la troisième vignette V_3					$Z(m,3)$
Valeur de $C(m)$	$C(m) = 4$				
$C(i) < C(m)$: l'individu i est en meilleure santé subjective que l'individu m pour la dimension de santé considérée.					

vignette-étalon, on doit toutefois leur adjoindre une hypothèse supplémentaire dans l'optique d'une *correction* de l'effet *DIF*. Notons k le nombre total de vignettes ($k > 1$).

H4. *Hypothèse d'ordre privilégié* : les k vignettes sont strictement ordonnées pour la dimension de santé considérée, selon $V_1 < V_2 < \dots < V_k$ par ordre de gravité strictement croissant.

Ainsi, pour la douleur physique dans l'enquête *Share*, nous ferons l'hypothèse que $V_1 < V_2 < V_3$ (pour les définitions des vignettes V_1 à V_3 , cf. encadré 2). Nous verrons que cet ordre, qui peut sembler assez naturel, n'est pas partagé par tous les répondants.

L'ordre strict mentionné dans l'hypothèse H4 sera appelé *ordre privilégié*. Pour éviter une trop grande instabilité des résultats obtenus, il est souhaitable qu'il y ait consensus des individus enquêtés, des analystes, d'experts ou des utilisateurs de l'enquête sur l'ordre privilégié des vignettes ; en l'absence d'un tel consensus, on doit faire face à de très difficiles problèmes d'agrégation des préférences, classiques en sciences politiques et économiques. *Cet aspect, peu abordé dans la littérature sur les vignettes-étalons, nous semble être une limitation importante de la méthodologie proposée, qui peut nuire à la crédibilité de l'analyse, en particulier dès lors que le nombre de vignettes est trop élevé.* En effet, nous verrons que la procédure proposée pour corriger l'effet *DIF* repose de façon essentielle sur le fait que les évaluations des vignettes par les répondants soient cohérentes avec l'hypothèse H4, ce qui requiert en particulier que des niveaux de santé suffisamment distincts soient associés aux différentes vignettes. Il y a donc un arbitrage à effectuer, lors de la conception du questionnaire, entre le nombre de vignettes à inclure et une différenciation suffisante de celles-ci.

Les hypothèses ci-dessus décrivent des situations idéales ; on peut donc difficilement s'attendre à ce qu'elles soient vérifiées exactement dans la pratique. Une validité approximative est toutefois souvent suffisante et divers arguments en faveur de celle-ci sont fournis par Salomon *et al.* (2000 et 2002), King *et al.* (2004) et King et Wand (2007). On mesure l'importance de la conception des vignettes et des modalités d'enquête et la nécessité de *tester* les vignettes avant toute utilisation à grande échelle ; diverses recommandations pratiques sont fournies par Salomon *et al.* (2000). De plus, la plupart des recommandations données dans la littérature

pour la conception d'instruments de mesure subjectives restent valables ici (Falissard, 2001, par exemple). On verra par ailleurs comment étudier la pertinence des hypothèses de travail à l'aide d'outils statistiques et, de façon connexe, comment prendre en compte les répondants dont les évaluations des vignettes ne sont pas cohérentes avec l'hypothèse H4, lors de la correction d'un effet *DIF*.

Détecter un problème de comparabilité

Les vignettes-étalons peuvent être utilisées pour *détecter* un effet *DIF* entre individus ou entre groupes d'individus. En effet, *sous les hypothèses* H1, H2, H3, il suffit, pour mettre en évidence un effet *DIF* et donc un problème de comparabilité des auto-évaluations entre deux répondants, de comparer les modalités de réponse utilisées pour l'évaluation des différentes vignettes. *Si deux individus répondent différemment à la question associée à une même vignette, les hypothèses d'équivalence des vignettes et d'équivalence des modalités de réponse impliquent qu'ils utilisent différemment les modalités de réponse lors de l'auto-évaluation ; leurs santé subjective ne peuvent alors être directement comparées à partir de leurs réponses à la question d'auto-évaluation.*

De façon analogue, *sous les hypothèses* H1, H2 et H3, il suffit, pour mettre en évidence un effet *DIF* entre sous-populations, de comparer les fréquences relatives d'utilisation des différentes modalités de réponse utilisées à l'issue de l'évaluation des différentes vignettes dans chaque sous-population.

Les réponses obtenues pour chacune des vignettes associées à la question d'auto-évaluation portant sur la douleur physique dans le questionnaire *Share* permettent d'illustrer cette utilisation des vignettes-étalons (cf. graphiques I à III). *À notre sens, celle-ci suffit à elle seule à justifier l'inclusion de vignettes-étalons dans des questionnaires comprenant des questions de santé subjective.*

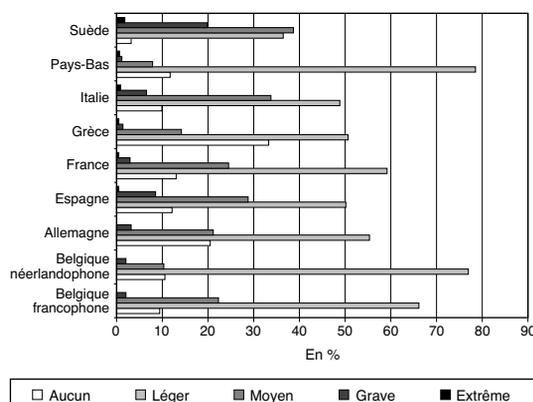
Ainsi, il existe une différence importante d'utilisation des modalités de réponse à l'évaluation du cas décrit par la vignette V_1 , entre l'échantillon suédois d'une part et les échantillons néerlandais, belge néerlandophone et grec d'autre part (cf. graphique I).

Une différence importante apparaît dans le cas de la vignette V_2 , entre l'échantillon suédois

d'une part et les autres échantillons d'autre part (cf. graphique II). La même différence est observée dans le cas de la vignette V_3 (cf. graphique III).

Graphique I
Première vignette associée à la douleur physique

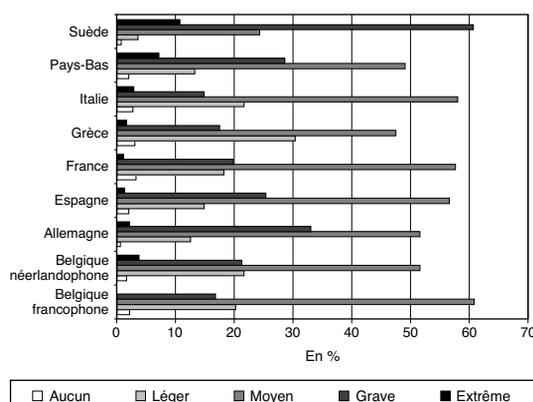
La répartition par échantillon des modalités de réponse



Lecture : répartition, en cinq modalités, des réponses à la question de la Première vignette associée au domaine de santé douleur physique dans le cadre de l'enquête Share 2004. « Paul a un mal de tête une fois par mois qui diminue après qu'il ait pris un cachet. Pendant qu'il a mal à la tête, il peut mener ses activités quotidiennes. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Paul a-t-il éprouvé ? » (cf. encadré 3). Champ : ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ». Source : enquête Share, 2004.

Graphique II
Deuxième vignette associée à la douleur physique

La répartition par échantillon des modalités de réponse



Lecture : répartition, en cinq modalités, des réponses à la question de la Deuxième vignette associée à la dimension de santé douleur physique dans le cadre de l'enquête Share 2004. « Henri a mal dans tout son bras droit et son poignet pendant sa journée de travail. Cela est partiellement atténué la soirée lorsqu'il ne travaille plus devant l'ordinateur. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Henri a-t-il éprouvé ? » (cf. encadré 3). Champ : ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ». Source : enquête Share, 2004.

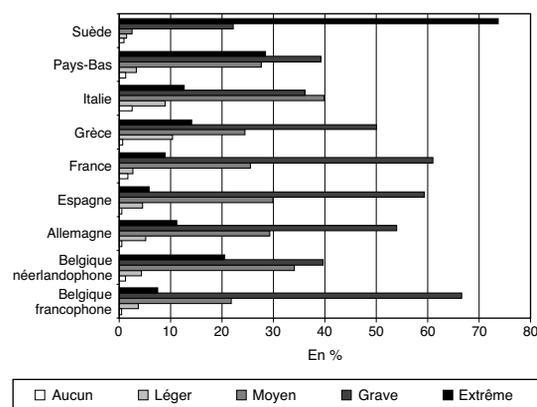
Ces constats suggèrent très fortement l'existence d'un effet *DIF* entre échantillons et donc un problème de comparabilité des auto-évaluations obtenues, notamment entre l'échantillon suédois et les autres échantillons. La crédibilité de cette conclusion repose essentiellement sur celle des hypothèses d'équivalence des vignettes et d'équivalence des modalités de réponse. L'hypothèse d'ordre privilégié H4 n'est, quant à elle, pas utilisée à ce stade de l'analyse.

Corriger les auto-évaluations pour améliorer la comparabilité

Une conséquence importante des hypothèses d'équivalence des vignettes, d'équivalence des modalités de réponse et d'unidimensionnalité est que le *positionnement de l'auto-évaluation d'un répondant par rapport à ses évaluations des vignettes associées* n'est pas affecté par l'effet *DIF* et peut donc être *déduit* de ses réponses. Ce positionnement peut alors être utilisé pour effectuer des comparaisons entre individus. En effet, un individu évaluant son état de santé avec une modalité supérieure (en gravité) à celle utilisée pour une vignette V_j donnée est en moins bonne santé subjective qu'un individu dont les évaluations des vignettes sont dans le même ordre que celles de l'individu précédent

Graphique III
Troisième vignette associée à la douleur physique

La répartition par échantillon des modalités de réponse



Lecture : répartition, en cinq modalités, des réponses à la question de la Troisième vignette associée au domaine de santé douleur physique dans le cadre de l'enquête Share 2004. « Charles a mal aux genoux, aux coudes, aux poignets et aux doigts, et la douleur est presque continuellement présente. Bien que les médicaments aident, il ne se sent pas bien lorsqu'il se déplace, qu'il tient ou soulève quelque chose. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Charles a-t-il éprouvé ? » (cf. encadré 3). Champ : ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ». Source : enquête Share, 2004.

et qui évalue son état de santé avec une modalité inférieure (en gravité) à celle utilisée pour cette même vignette V_j .

Dans la situation idéale où l'ordre sur les vignettes qui est déduit d'une comparaison de leurs évaluations est le même pour tous les

Encadré 4

**DÉFINITION D'UN INTERVALLE D'INCERTITUDE
POUR L'AUTO-ÉVALUATION CORRIGÉE DE L'EFFET DIF
- DIMENSION DE SANTÉ « DOULEUR PHYSIQUE » - SHARE 2004**

À la suite de King et Wand (2007), nous pouvons en général obtenir, pour chaque individu i dont les réponses ne sont pas cohérentes, un ensemble *non vide* $G(i)$ de vignettes-étalons susceptibles d'être équivalentes à la situation de l'individu i pour la dimension de santé considérée.

Convenons de noter, $r(i, -)$ le rang, dans la série des vignettes ordonnées selon l'ordre privilégié de la vignette élément de $G(i)$ correspondant à l'état de santé le moins grave et $r(i, +)$ le rang, dans la série des vignettes ordonnée selon l'ordre privilégié, de la vignette élément de $G(i)$ correspondant à l'état de santé le plus grave.

On pose alors $C(i, -) = 2r(i, -)$ et $C(i, +) = 2r(i, +)$.

$C(i, -)$ correspond à la valeur de C qui serait associée à un individu i dont les réponses seraient cohérentes et dont l'état de santé pour la dimension considérée correspondrait à peu près à celui décrit dans la vignette de rang $r(i, -)$.

$C(i, +)$ correspond à la valeur de C qui serait associée à un individu i dont les réponses seraient cohérentes et dont l'état de santé pour la dimension considérée cor-

respondrait à peu près à celui décrit dans la vignette de rang $r(i, +)$.

Pour un individu i dont les réponses sont cohérentes, on pose $C(i, -) = C(i, +) = C(i)$.

À titre d'illustration, soit i un répondant tel que $Y(i) = \text{« moyen »}$, $Z(i,1) = \text{« moyen »}$, $Z(i,2) = \text{« léger »}$, $Z(i,3) = \text{« grave »}$. Comme $Z(i,2) < Z(i,1)$, les réponses de l'individu ne sont pas cohérentes. La procédure présentée par King et Wand (2007) permet d'obtenir $G(i) = \{V_1, V_2\}$, c'est-à-dire que V_1 et V_2 sont toutes deux susceptibles de correspondre à la situation de l'individu i en termes de niveau de douleur physique. On en déduit $r(i, -) = 1$, $r(i, +) = 2$, d'où $C(i, -) = 2$ et $C(i, +) = 4$.

Une présentation détaillée et illustrée de la procédure d'obtention de $G(i)$ est donnée dans King et Wand (2007) et une implémentation informatique de celle-ci est disponible dans le package « anchors » du logiciel R développé par King et Wand (2005), que nous avons utilisé et adapté aux besoins de notre étude. King et Wand (2007) proposent également une méthode permettant de traiter le cas des individus pour lesquels l'ensemble $G(i)$ est vide.

Tableau A
Deux cas d'incohérence

Premier cas : cas où la procédure de King et Wang conduit à ne pas retenir la vignette V_3

	Aucun	Léger	Moyen	Grave	Extrême
Réponse à la question de l'enquête Share			$Y(i)$		
Réponse à la première vignette V_1			$Z(i,1)$		
Réponse à la deuxième vignette V_2		$Z(i,2)$			
Réponse à la troisième vignette V_3				$Z(i,3)$	
Valeur de $r(i, -)$ (rang de V_1)			1		
Valeur de $r(i, +)$ (rang de V_2)		2			
Valeur de $C(i, -)$ (évaluation « optimiste »)			2		
Valeur de $C(i, +)$ (évaluation « pessimiste »)		4			

Deuxième cas : cas où la procédure de King et Wang conduit à ne pas retenir la vignette V_1

	Aucun	Léger	Moyen	Grave	Extrême
Réponse à la question de l'enquête Share				$Y(m)$	
Réponse à la première vignette V_1			$Z(m,1)$		
Réponse à la deuxième vignette V_2					$Z(m,2)$
Réponse à la troisième vignette V_3		$Z(m,3)$			
Valeur de $r(m, -)$ (rang de V_1)					2
Valeur de $r(m, +)$ (rang de V_2)		3			
Valeur de $C(m, -)$ (évaluation « optimiste »)					4
Valeur de $C(m, +)$ (évaluation « pessimiste »)		6			

répondants, il est possible de comparer l'état de santé de deux répondants quelconques. Lorsque c'est le cas, on peut naturellement adopter cet ordre commun comme ordre privilégié dans l'hypothèse H4. Dans le cas plus général, et plus réaliste, où l'ordre privilégié n'est pas un ordre partagé par l'ensemble des répondants, on introduit la notion.

On introduit la notion de *consistance* : par souci de concision, les réponses d'un individu enquêté seront qualifiées simplement de *cohérentes* si l'évaluation qu'il fait des vignettes est cohérente avec l'hypothèse H4 c'est-à-dire si elle respecte l'ordre privilégié.

Un individu dont les réponses sont *cohérentes* et qui évalue son état de santé à l'aide d'une modalité strictement supérieure (en gravité) à celle utilisée pour une vignette donnée, est en moins bonne santé subjective qu'un individu dont les réponses sont *cohérentes* et qui évalue son propre état de santé à l'aide d'une modalité strictement inférieure (en gravité) à celle utilisée pour cette même vignette, et ce *indépendamment de la population à laquelle il appartient*. Dans ce type de comparaisons, les vignettes sont utilisées comme des *étalons* ou *points d'ancrage* permettant de comparer les auto-évaluations. Afin de systématiser cette approche, King *et al.* (2004, 2007) proposent de définir, sous les hypothèses H1 à H4 et pour les répondants dont les réponses sont cohérentes, une variable numérique ordinale C traduisant le positionnement de l'auto-évaluation par rapport aux évaluations des vignettes (cf. encadré 3). La variable ainsi obtenue peut être interprétée comme une version de l'auto-évaluation *corrigée non paramétriquement* de l'effet *DIF*, c'est-à-dire sans recours à une modélisation aléatoire paramétrique du comportement de réponse des individus (King *et al.*, 2004 et 2007).

La variable C peut être utilisée de façon très simple pour comparer l'état de santé de deux répondants dont les réponses sont cohérentes. Par exemple, si $C(i) < C(j)$, il est cohérent d'affirmer que l'individu i est en meilleure santé subjective que l'individu j pour la dimension de santé considérée. Une limite évidente de cette approche est que C n'est définie que pour les répondants dont les réponses sont cohérentes et, par conséquent, ne permet pas d'effectuer des comparaisons d'état de santé subjectif entre sous-populations comprenant des répondants dont les réponses ne sont pas cohérentes. Afin de pallier ces inconvénients, King et Wand (2007) proposent une approche plus générale, que nous adaptons de

façon originale dans ce travail. Il s'agit d'une solution qui n'est pas exempte de critique mais qui permet à tout le moins de mettre en évidence et de quantifier l'influence des problèmes de consistance sur l'issue de l'analyse.

L'idée développée par King et Wand (2007) consiste à identifier, pour chaque individu i dont les réponses ne sont pas cohérentes, un ensemble $G(i)$ de vignettes susceptibles d'être équivalentes à la situation de l'individu i pour la dimension de santé considérée (cf. encadré 4). Une présentation détaillée et illustrée de la procédure d'obtention de $G(i)$ est donnée dans King et Wand (2007) et une implémentation informatique de celle-ci est disponible dans le *package* « *anchors* » du logiciel *R* développé par King et Wand (2005), que nous avons utilisé et adapté aux besoins de notre étude.

Nous utilisons les ensembles $G(\cdot)$ et la variable C pour définir, pour chaque répondant i , un couple de valeurs $(C(i, -), C(i, +))$ décrivant le résultat de son auto-évaluation relativement aux différentes vignettes.

L'introduction des variables $C(\cdot, -)$ et $C(\cdot, +)$ est à notre connaissance, spécifique à notre travail. Nous proposons d'interpréter $C(\cdot, -)$ comme une évaluation *optimiste* de la santé subjective de l'individu i et $C(i, +)$ comme une évaluation *pessimiste* de la santé subjective de l'individu i . On dispose ainsi d'un *intervalle ordinal d'incertitude* sur l'état de santé subjectif de *chaque individu*, que ses réponses soient cohérentes ou non.

Les variables $C(\cdot, -)$ et $C(\cdot, +)$ ainsi définies peuvent alors être utilisées pour obtenir, pour chaque population et chaque vignette V_j , un *intervalle d'incertitude* pour la proportion d'individus évaluant leur état de santé équivalent ou meilleur que celui décrit dans la vignette V_j et un *intervalle d'incertitude* pour la proportion d'individus évaluant leur état de santé équivalent ou moins bon que celui décrit dans la vignette V_j (cf. encadré 5). Les intervalles obtenus pour diverses sous-populations peuvent alors être comparés à l'aide d'outils exploratoires, notamment graphiques.

Il est, à ce stade, important de noter que cette construction et son interprétation sont relatives à l'ordre privilégié et qu'elles peuvent donc être critiquées sur cette base. Nous verrons toutefois que le fait d'obtenir des intervalles d'incertitude permet de *quantifier l'importance des problèmes de non-consistance lors de comparaisons entre sous-populations* et donc de mettre en évidence l'ampleur des écarts potentiels aux hypo-

thèses de travail, ce qui permettra de relativiser certains résultats obtenus.

Comparer l'état de santé de sous-populations : une application de la méthode aux échantillons *Share*

Illustrons l'utilisation de ces intervalles d'incertitude par une comparaison des échantillons *Share* par origine, pour la dimension de santé douleur physique.

L'échantillon suédois se distingue clairement des autres échantillons par une proportion plus élevée d'individus évaluant leur état de santé équivalent ou meilleur que celui décrit dans la vignette V_1 et par une incertitude très faible. L'échantillon néerlandais se distingue également de la plupart des autres échantillons, mais pas de l'échantillon italien, caractérisé par une incertitude très importante, donc par des problèmes de consistance importants (cf. graphique IV).

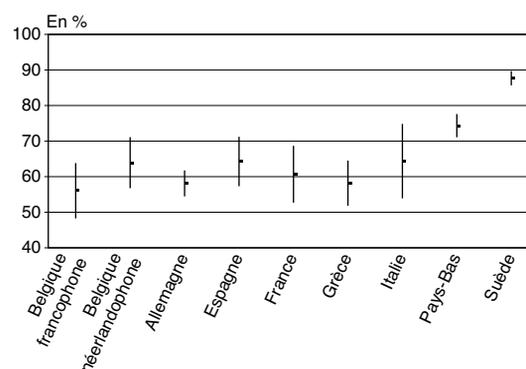
L'examen des proportions d'individus évaluant leur état de santé équivalent ou meilleur que celui décrit dans la vignette V_2 (cf. encadré 2) pour ce qui concerne la douleur physique conduit à des conclusions analogues (cf. graphique V).

On remarque que l'échantillon suédois se distingue par la proportion la plus élevée d'individus évaluant leur état de santé équivalent ou meilleur que celui décrit dans la vignette V_2 et par une incertitude très faible. On peut également avancer que la proportion d'individus évaluant leur état de santé équivalent ou meilleur que celui décrit dans la vignette V_2 est plus élevée au sein de l'échantillon néerlandais qu'au sein de l'échantillon grec. Par contre, aucune

des autres comparaisons entre échantillons ne permet de suggérer de différence significative, les différents intervalles d'incertitude se recouvrant partiellement les uns les autres. Dans le cas de l'échantillon italien, cette conclusion doit encore être relativisée par l'importance des problèmes de consistance.

La proportion d'individus évaluant leur état de santé équivalent ou moins bon que celui décrit

Graphique IV
Proportion par échantillon de ceux qui évaluent, du point de vue de la douleur physique, leur état de santé équivalent ou meilleur que l'appréciation qui est la leur du cas décrit dans la première vignette



Lecture : en Grèce, 58 % des répondants évaluent leur état de santé meilleur (du point de vue de la douleur physique) que l'évaluation qu'ils font du cas présenté dans la première vignette associée au domaine de santé douleur physique dans le cadre de l'enquête *Share* 2004 : « Paul a un mal de tête une fois par mois qui diminue après qu'il ait pris un cachet. Pendant qu'il a mal à la tête, il peut mener ses activités quotidiennes. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Paul a-t-il éprouvé ? » (cf. encadré 3). L'existence de cas de non-cohérence conduit à définir un intervalle d'incertitude allant de 53 % à 64 % (cf. encadrés 5 et 6).

Champ : ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ».

Source : enquête *Share*, 2004.

Encadré 5

DÉFINITION DES INTERVALLES D'INCERTITUDES (PROPORTIONS)

Notons $F(x, -)$ la proportion d'individus pour lesquels $C(\circ, -)$ prend une valeur inférieure ou égale à x et $F(x, +)$ la proportion d'individus pour lesquels $C(\circ, +)$ prend une valeur inférieure ou égale à x dans l'une des sous-populations considérées. Alors, pour toute vignette V_j , l'intervalle $I(j, -) = [F(2j, +), F(2j, -)]$ est un *intervalle d'incertitude* pour la proportion d'individus dont l'état de santé subjectif est équivalent ou meilleur que celui décrit dans la vignette V_j .

Notons $G(x, -)$ la proportion d'individus pour lesquels $C(\circ, -)$ prend une valeur supérieure ou égale à x et $G(x, +)$ la proportion d'individus pour lesquels $C(\circ, +)$ prend une valeur supérieure ou égale à x dans l'une des sous-populations considérées. Alors, pour toute vignette V_j , l'intervalle $I(j, +) = [G(2j, -), G(2j, +)]$ est un *intervalle d'incertitude* pour la proportion d'individus dont l'état de santé subjectif est équivalent ou moins bon que celui décrit dans la vignette V_j .

dans la vignette V_3 (cf. encadré 2) est moindre au sein de l'échantillon suédois qu'au sein des autres échantillons et l'incertitude est, là encore, très faible (cf. graphique VI). D'autre part, les autres échantillons ne se distinguent pas clairement les uns des autres pour ce critère, même si, dans le cas de l'échantillon italien, cette conclusion doit être relativisée par l'importance des problèmes de consistance.

Les remarques précédentes amènent naturellement à conclure, au vu des données et sous les hypothèses faites, que l'échantillon suédois est globalement en meilleure santé subjective que les autres échantillons en termes de douleur physique et qu'en dehors d'éventuels résultats partiels, concernant les échantillons néerlandais et grec notamment, il n'y a aucune différence significative d'état de santé subjectif global entre les autres échantillons pour cette même dimension.

Ces résultats diffèrent significativement de ceux obtenus en se basant uniquement sur les auto-évaluations non corrigées de l'effet *DIF* et conduisent à relativiser les comparaisons basées sur celles-ci. À titre d'illustration, considérons les réponses à la question d'auto-évaluation

obtenues pour échantillons suédois et néerlandais (cf. tableau).

L'utilisation des auto-évaluations brutes amène naturellement à postuler que les échantillons suédois et néerlandais ne se distinguent pas vrai-

Tableau
Fréquences cumulées des auto-évaluations de la douleur physique

En %

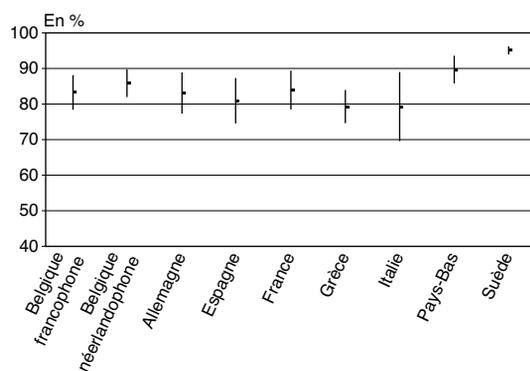
Réponse	Pays-Bas	Suède
Aucun	37,5	52,5
Léger ou moins	81,0	81,6
Moyen ou moins	94,8	92,1
Moyen ou plus	19,0	18,4
Grave ou plus	5,2	7,9
Extrême	1,7	1,0

Lecture : aux Pays-Bas, 94,8 % des personnes interrogées ont donné une des trois réponses « Aucun », « Léger », « Moyen », 5,2 % répondant « Grave » ou « Extrême » à la question : « Dans l'ensemble, au cours des 30 derniers jours, quel niveau de douleurs physiques avez-vous ressenti ? » (cf. encadré 2).

Champ : répondants néerlandais et suédois à la question d'auto-évaluation et aux questions de vignettes associée à la dimension de santé « douleur physique ».

Source : enquête Share, 2004.

Graphique V
Proportion par échantillon de ceux qui évaluent, du point de vue de la douleur physique, leur état de santé équivalent ou meilleur que l'appréciation qui est la leur du cas décrit dans la deuxième vignette



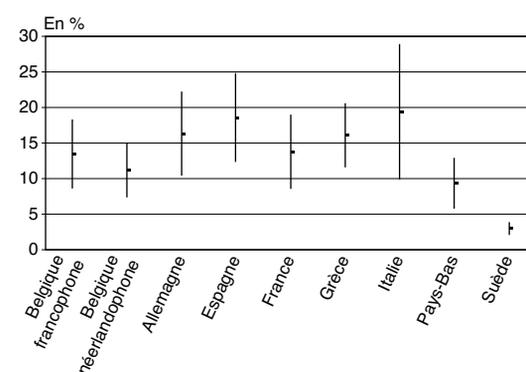
Lecture : en Grèce, 79 % des répondants évaluent leur état de santé meilleur (du point de vue de la douleur physique) que l'évaluation qu'ils font du cas présenté dans la deuxième vignette associée au domaine de santé douleur physique dans le cadre de l'enquête Share 2004 : « Henri a mal dans tout son bras droit et son poignet pendant sa journée de travail. Cela est partiellement atténué la soirée lorsqu'il ne travaille plus devant l'ordinateur. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Henri a-t-il éprouvé ? » (cf. encadré 3).

L'existence de cas de non-cohérence conduit à définir un intervalle d'incertitude allant de 75 % à 84 % (cf. encadrés 5 et 6).

Champ : Ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ».

Source : enquête Share, 2004.

Graphique VI
Proportion par échantillon de ceux qui évaluent, du point de vue de la douleur physique, leur état de santé équivalent ou moins bon que l'appréciation qui est la leur du cas décrit dans la troisième vignette



Lecture : en Grèce, 17 % des répondants évaluent leur état de santé équivalent ou moins bon (du point de vue de la douleur physique) que l'évaluation qu'ils font du cas présenté dans la première vignette associée au domaine de santé douleur physique dans le cadre de l'enquête Share 2004 : « Charles a mal aux genoux, aux coudes, aux poignets et aux doigts, et la douleur est presque continuellement présente. Bien que les médicaments aident, il ne se sent pas bien lorsqu'il se déplace, qu'il tient ou soulève quelque chose. En général, au cours des 30 derniers jours, quel niveau de douleurs physiques Charles a-t-il éprouvé ? » (cf. encadré 3).

L'existence de cas de non-cohérence conduit à définir un intervalle d'incertitude allant de 12 % à 21 % (cf. encadré 5 et 6).

Champ : ensemble des individus ayant répondu à la question d'auto-évaluation et aux questions de vignettes associées pour la dimension de santé « douleur physique ».

Source : enquête Share, 2004.

ment en terme de santé subjective, voire qu'il y a davantage d'individus subjectivement en mauvaise santé, pour ce qui est de la douleur physique, au sein de l'échantillon suédois qu'au sein de l'échantillon néerlandais, ce qui contraste fortement avec les résultats de l'analyse basée sur les auto-évaluations corrigées de l'effet *DIF potentiel*. L'utilisation très différente des modalités de réponse lors de l'évaluation les vignettes au sein des échantillons suédois et néerlandais

(cf. graphiques I à III) suggère fortement l'existence d'un effet *DIF* et d'une différence en termes d'attentes ou de norme pour la dimension de santé considérée entre ces deux échantillons. Ce n'est bien entendu pas la seule explication possible, mais les résultats précédents incitent à aborder avec circonspection toute comparaison de la santé subjective des échantillons suédois et néerlandais à partir des auto-évaluations non corrigées.

Pour la réalisation de cette recherche, nous nous sommes servis des données issues de la première version (Release 1) de la vague 1 de l'enquête Share réalisée en 2004. Celle-ci est préliminaire et pourrait contenir des erreurs qui seront corrigées dans des versions ultérieures. La collecte des données de Share a été principalement financée par le 5^e Programme Cadre de la Communauté Européenne (Projet QLK6-CT-2001-00360 sur le thème de la qualité de la vie). Des fonds proviennent également du US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 et OGHA 04-064). La collecte des données a été financée par des institutions nationales en Autriche (Austrian Science Foundation, FWF), en Belgique (Politique Scientifique Fédérale) et en Suisse (BBW/OFES/UFES). Pour la France, elle a bénéficié de financements complémentaires apportés par la Cnav, la Cnam, le Cor, la Drees, la Dares, la Caisse des Dépôts et Consignations et le Commissariat Général du Plan. Les données et la méthodologie de l'enquête sont respectivement présentées dans Börsch-Supan et al. (2005) et Börsch-Supan et Jürges (2005). □

BIBLIOGRAPHIE

Alexander C.S. et Becker H.J. (1978), « The Use of Vignettes in Survey Research », *The Public Opinion Quarterly*, vol. 42, n° 1, pp. 93-104.

Börsch-Supan A., Brugiavini A., Jürges H., Mackenbach J., Siegrist J. et Weber G. (2005), *Health, Ageing and Retirement in Europe. First Results from the Survey of Health, Ageing and Retirement in Europe*, Mannheim Research Institute for the Economics of Aging, MEA, Mannheim.

Börsch-Supan A. et Jürges H. (éds.) (2005), *The Survey of Health, Ageing and Retirement in Europe - Methodology*, Mannheim : MEA.

Falissard B. (2001), *Mesurer la subjectivité en Santé*, Masson, Paris.

King G., Murray C.J.L., Salomon J.A. et Tandon A. (2004), « Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research (corrected version) », *American Political Science Review*, vol. 98, n° 1, pp. 191-207.

King G. et Wand J. (2005), *Anchors : Software for Anchoring Vignette Data*. Documentation R, disponible à l'adresse <http://GKing.Harvard.edu/vign/>.

King G. et Wand J. (2007), « Comparing Incomparable Survey Responses : New Tools for Anchoring Vignettes », *Political Analysis*, à paraître.

Murray C.J.L. et Chen L.C. (1992), « Understanding Morbidity Change », *Population and Development Reviews*, vol. 18, n° 3, pp. 481-503.

Nunnally J.C. et Bernstein I.H. (1994), *Psychometric Theory*, McGraw-Hill.

Salomon J.A., Murray C.J.L. et Tandon A. (2002), « Using Vignettes to Improve Cross-Population Comparability of Health Surveys : Concepts, Design, and Evaluation Techniques », World Health Organization, *GPE Discussion Paper*, n° 42.J.

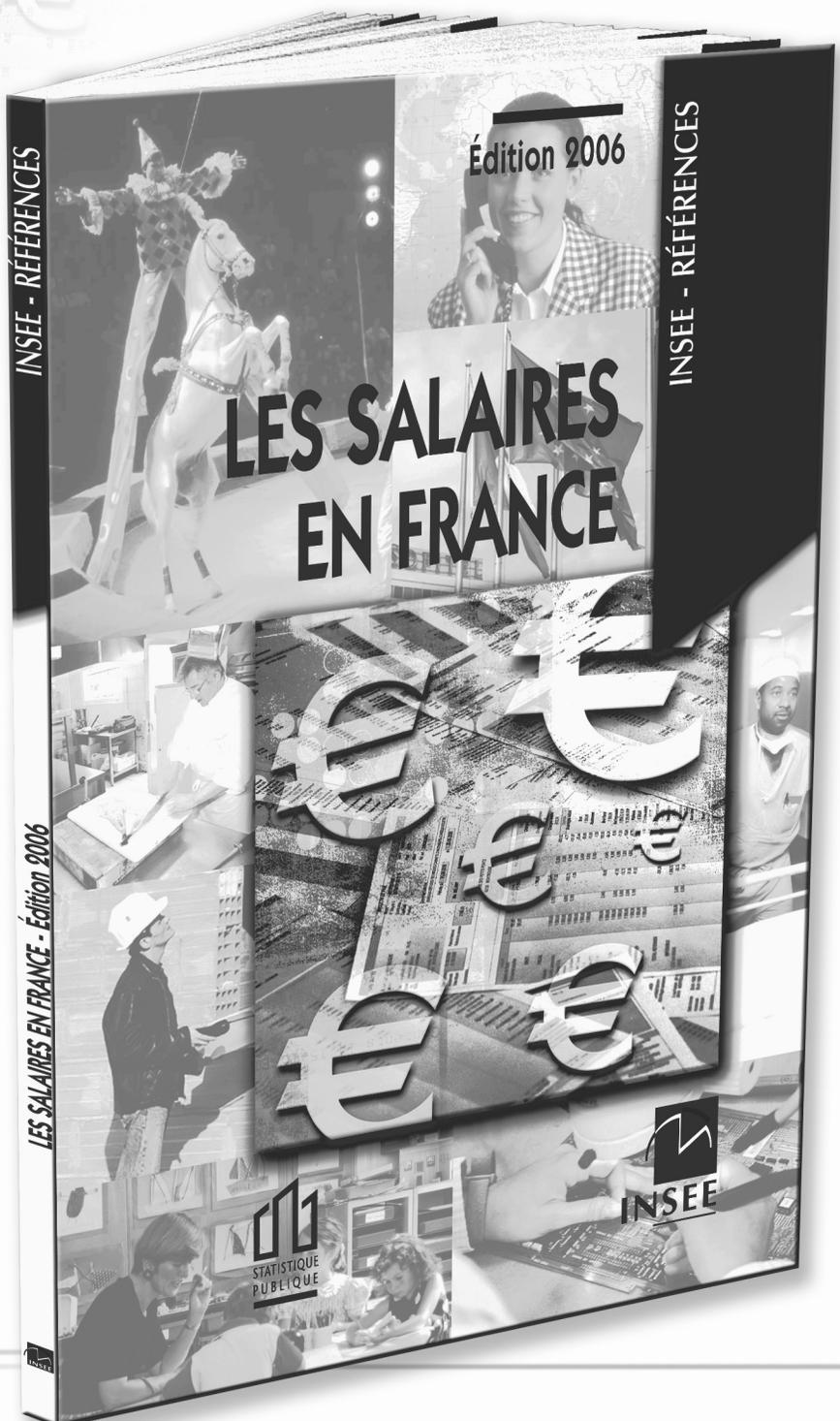
Salomon J.A., Tandon A. et Murray C.J.L. (2004), « Comparability of Self Rated Health : Cross Sectional Multi-Country Survey Using Anchoring Vignettes », *British Medical Journal*, vol. 328, n° 7434, pp. 258-260.

Sen A. (2002), « Health : Perception versus Observation », *British Medical Journal*, vol. 324, n° 7342, pp. 860-861.

World Health Organization (1980), *International Classification of Impairments, Disabilities, and Handicaps : a Manual of Classification Relating to the Consequences of Disease*, World Health Organization, Geneva.

World Health Organization (1999), *ICIDH-2 : International Classification of Functioning and Disability*, Beta-2 draft, short version. World Health Organization, Geneva.

Ce que gagnent les Français...



- Trois dossiers sur les disparités salariales et leurs déterminants, l'épargne salariale et une analyse en Europe.
- Un zoom thématique sur les principaux secteurs économiques (industrie, construction, tertiaire...), selon l'âge, le sexe...

En vente en librairie,
à l'Insee et sur www.insee.fr

15 € - Collection Insee-Références


INSEE