A contribution to the analysis of inequalities in health in France using indicators of self-assessed health

Sandy Tubeuf

▶ **To cite this version:**

Sandy Tubeuf. A contribution to the analysis of inequalities in health in France using indicators of self-assessed health. Economics and Finance. Université de la Méditerranée - Aix-Marseille II, 2008. English. NNT : . tel-00288887

HAL Id: tel-00288887

https://theses.hal.science/tel-00288887

Submitted on 19 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Aix-Marseille II

UFR Sciences Economiques et de Gestion

Une contribution à l'étude des inégalités de santé en France à travers des indicateurs de santé auto-évalués

par

Sandy TUBEUF

Thèse pour le Doctorat en
Sciences Economiques
(Arrêté du 30 mars 1992)

Co-directeurs de Recherche :

Lise ROCHAIX, HAS - Professeur à l'Université Aix-Marseille II
Alain TRANNOY, Directeur d'études à l'EHESS et au GREQAM

Jury:

Chantal CASES, Directrice de l'IRDES
Brigitte DORMONT, Professeur à l'Université Paris Dauphine, Rapporteur
Andrew JONES, Professeur à l'Université de York, Rapporteur
Jérôme WITTWER, Professeur à l'Université Paris Dauphine

Présentée et soutenue publiquement le 11 Février 2008

L'Université d'Aix-Marseille II n'entend donner aucune approbation ou improbation aux opinions émises dans cette thèse : ces opinions doivent être considérées comme propres à leurs auteurs.

En application de la loi Toubon du 4 août 1994, le Professeur Jacques Derrien, Vice-Président à la Recherche de l'Université de la Méditerranée II a rappelé, dans sa lettre du 26 avril 2007, au sujet de la rédaction des thèses que “s'agissant d'une thèse hors co-tutelle, l'Université admet une présentation de la couverture entièrement en français, une introduction en français, une conclusion en français. Tout le reste peut être organisé soit en bilingue soit en anglais”.

Conformément à la loi, les chapitres introductifs et conclusifs ont donc été rédigés en français. Les cinq chapitres présentant la recherche sont rédigés en anglais, seulement.

Résumé

Une contribution à l'étude des inégalités de santé en France à travers des indicateurs de santé auto-évalués

par

Sandy TUBEUF

Université Aix-Marseille II

Lise ROCHAIX et Alain TRANNOY, Co-directeurs de Recherche

Cette thèse s'inscrit dans le champ de la mesure et de l'explication de la santé dans un contexte d'analyse des inégalités de santé.

Un premier chapitre considère les indicateurs de santé couramment utilisés dans les travaux empiriques et revient sur le débat de l'utilisation de la santé auto-évaluée. Il souligne la pertinence des raffinements méthodologiques de la mesure de la santé proposés dans la littérature internationale jusqu'ici non appliqués à la France.

Un second chapitre propose une méthodologie originale de mesure de la santé. La construction s'appuie sur une donnée d'état de santé individuel jugée moins subjective, à savoir le nombre de maladies et leur degré de sévérité et considère des variables collectées classiquement dans les enquêtes sur la santé.

Un troisième chapitre décrit les outils de la dominance stochastique et les indices couramment utilisés dans l'analyse des inégalités dans un cadre appliqué à la santé.

Le quatrième chapitre procède à l'analyse des inégalités sociales de santé en France en 2004, puis au cours de la période 1998-2004. Il met en évidence des inégalités sociales de santé en faveur des groupes sociaux les plus élevés. Ces inégalités ont cependant diminué entre 1998 et 2004, du fait d'une plus faible élasticité de la santé avec le revenu et d'une diminution de l'inégale répartition du revenu au sein des groupes sociaux. De plus, l'analyse menée sur différentes mesures de santé met en évidence une influence sur l'amplitude des inégalités, du nombre de catégories de la variable discrète de santé et de la distribution de santé choisie pour la cardinaliser.

Le cinquième chapitre s'intéresse à l'influence sur l'état de santé à l'âge adulte, du milieu social d'origine et de la longévité relative des parents par rapport à leur cohorte de naissance en empruntant trois approches. La première approche met en évidence le fait que les distributions d'état de santé des personnes nées d'un père ou d'une mère appartenant aux catégories sociales supérieures dominent significativement celles des personnes ayant des parents issus de catégories sociales inférieures. L'approche paramétrique confirme un effet de la profession de chacun des parents sur l'état de santé à l'âge adulte. Elle montre, de plus, que l'état de santé dépend significativement de la longévité de chacun des parents. Enfin, l'approche par indices de concentration met en évidence une inégalité des chances de santé en faveur des individus dont les parents ont connu une forte longévité puis une inégalité de santé en faveur des individus issus de milieux plus favorisés. Le chapitre conclut alors qu'il existe des inégalités des chances en santé, en France.

Mots-clefs : Dominance stochastique du premier ordre - Indicateurs de santé - Inégalités - Inégalités des chances - Indice de concentration - Santé auto-évaluée

Classification JEL : C14 - C43 - D63 - I10 - I14

Abstract

A contribution to the analysis of inequalities in health in France using indicators of self-assessed health

by
Sandy TUBEUF

University Aix-Marseille II,

Lise ROCHAIX and Alain TRANNOY, Co-supervisors

This dissertation is devoted to two main topics: the measurement and the explanation of health within an analysis of inequalities in health status.

In chapter 1, we review health indicators which are commonly used within empirical analyses and take part in the debate on the use of self-assessed health. We show that there is a challenge to apply recent methodological refinements of the measurement of health proposed in the international literature, which at present, lacks in the French context.

Chapter 2 develops an innovative method of constructing a concrete measure of health by taking into account an individual health information being considered less subjective, namely the number of diseases and their respective severity level. Moreover, the construction considers several variables which are classically collected in health surveys.

Chapter 3 describes, within a framework applied to health, measurement tools such as stochastic dominance and inequality indices, which are commonly used for inequalities analyses.

In chapter 4, we analyse income-related inequalities in health in France in 2004 and in the period 1998-2004. The analysis shows income-related inequalities in health favouring socially advantaged groups. These inequalities have nevertheless decreased between 1998 and 2004, this reduction driven by a lower elasticity of health with income and lower inequalities in income over socioeconomic groups. The analysis being carried out with alternative measurements of health, inequalities in health appear to vary quantitatively with both the number of categories of self-assessed health and the distribution of health used to cardinalise this variable.

Chapter 5 investigates the explanation of health status in adulthood according to social and family background, as measured by the mother's and the father's socioeconomic status, as well as their respective longevity. Three methods are used. The first approach emphasises that distributions of health of individuals born to a father or a mother in higher socioeconomic statuses significantly dominate distributions of health of individuals born to a father or a mother in lower socioeconomic statuses. The parametric approach confirms the effect of social background on health in adulthood. Furthermore, it shows that health in adulthood can be explained significantly by parents' longevity. To conclude, concentration indices of inequalities describe inequalities favouring individuals whose parents had lived longer and inequalities in health favouring individuals born to socially advantaged families. Therefore, this chapter concludes that in France there is a family and social determinism in health.

Keywords: Concentration index - Health indicators - Inequalities - Inequalities of opportunity - Self-assessed health - Stochastic dominance at first-order

JEL codes : C14 - C43 - D63 - I10 - I14

A mes parents,

"Degitur hoc aevi quod cumquest ! Nonne videre
Nihil aliud sibi naturam latrare, nisi ut qui
Corpore seiunctus dolor absit, mensque fruatur
Iucundo sensu cura semota metuque ?
Ergo corpoream ad naturam pauca videmus
Esse opus omnino : quae demant cumque dolorem,
Delicias quoque uti multas substernere possint
Gratius inter dum, neque natura ipsa requirit"

"Où notre courte vie s'écoule ? Entendez-vous
Ce que crie la nature ? Elle veut pour le corps
L'absence de douleur, elle attend pour l'esprit
Un bien-être à l'abri des craintes et soucis.
On voit ainsi le corps se contenter de peu.
Tout ce qui de ses maux le peut débarrasser
Lui donne également des joies délicieuses."

Lucrèce (98-54)
De rerum natura - Liber II (v. 16-24)

Remerciements

Il me tient à coeur de dire merci à ceux qui ont concouru de près comme de loin à la réalisation de cette thèse.

En premier lieu, je souhaite exprimer toute ma gratitude envers Lise Rochaix et Alain Trannoy. Ils ont l'un comme l'autre éveillé ma réflexion à des pistes de recherches originales et dynamiques. Ils m'ont toujours encouragée à aller chercher plus loin et je leur suis reconnaissante pour le temps que nous avons passé ensemble. Leur double encadrement m'a donné la chance de goûter à une complémentarité d'encadrement dont je mesure aujourd'hui l'enrichissement tant au plan scientifique qu'au plan humain. Je les remercie sincèrement pour leur soutien continuel, leur aide inestimable et leur enthousiasme pour ces travaux.

Je tiens ensuite à remercier Brigitte Dormont et Andrew Jones qui me font l'honneur d'être rapporteurs de cette thèse. Mes travaux ont eu la chance de bénéficier au moment de la présoutenance de leurs relectures attentives et de leurs précieuses critiques qui m'ont permis d'améliorer substantiellement cette thèse.

En particulier, je remercie Brigitte Dormont pour ses conseils à l'occasion des Journées Doctorales en Economie de la Santé au tout début de ma thèse. J'ai aussi apprécié à plusieurs reprises son enthousiasme pour ces questions de la mesure de l'état de santé et ses encouragements à investir ce champ d'analyse.

Andrew Jones m'a entourée de sa bienveillance durant mes séjours à York et à Bergen, j'ai beaucoup appris à ses côtés et ce travail empirique a reçu ses précieux conseils économétriques.

Je remercie Jérôme Wittwer pour l'intérêt qu'il a manifesté pour ces questionnements sur les inégalités des chances en santé lors du séminaire Eurisco-Legos à l'université de Paris-Dauphine et j'apprécie sa présence dans ce jury de thèse.

Je remercie Chantal Cases d'avoir accepté de prendre part au jury de cette thèse. A travers elle, c'est tout l'Irdes que je tiens à remercier.

L'Irdes a été la maison chaleureuse où j'ai fait mes premiers pas dans la recherche en économie de la santé. J'ai beaucoup reçu tant du point de vue professionnel, que du point de vue intellectuel. Je remercie pour leurs “bons soins” durant quatre ans, non seulement Dominique Polton, Chantal Cases et Catherine Sermet, directrices et co-directrice successives de l'IRDES mais aussi l'ensemble de l'équipe de direction, la documentation, l'équipe ESPS, les informaticiens, les maquettistes et toute l'équipe de recherche.

Je remercie aussi sincèrement Andrew Jones, Nigel Rice et Peter Smith de m'avoir fait une place à l'Université de York à deux reprises durant cette thèse. Le Centre for Health Economics ainsi que le Health Econometrics and Data Group sont des endroits privilégiés pour la recherche et je suis reconnaissante de cette chance qui m'a été donnée d'y rester plusieurs mois.

Ces séjours à l'Université de York sont les fruits du réseau ECuity dans lequel j'ai eu la chance d'être impliquée. Dans ce contexte extraordinaire mêlant une grande qualité scientifique et un excellent encadrement de doctorants, j'ai pu généreusement bénéficier des commentaires avisés de chercheurs que je remercie. Je pense en particulier à Jan Erik Askilden, Terkel Christiansen, Owen O' Donnell, Eddy van Doorslaer, Brenda Gannon, Diana de Graeve, Unto Häkkinen, Xander Koolman, Maarten Lindeboom, Angel Nicolas Lopez, Tom van Ourti, Carol Propper, Martin Schellhorn ainsi qu'à mes "consoeurs" doctorantes européennes.

J'ai eu la chance durant ma thèse d'élargir mes connaissances au sein de cours doctoraux et de présenter mes travaux à diverses conférences. Parmi les personnes qui m'ont beaucoup apporté dans ces deux domaines, je souhaite vivement remercier Andrew Clark, Agnès Couffinhal, Fabrice Etilé, Marc Fleurbaey, Carine Franc, Pierre-Yves Geoffard, Michel Grignon, Alberto Holly, Florence Jusot, Myriam Khlat, Massimo Philippini, Aki Tsuchiya, Bruno Ventelou, les membres du comité scientifique de l'IRDES ainsi que les participants du réseau EPISOC, du séminaire Eurisco-Legos à l'université de Paris-Dauphine, des mardis de l'IRDES et des lundis de l'Institut National des Etudes Démographiques.

Pour nos échanges sur les indicateurs de qualité de vie, je remercie Ling-Hsiang Chuang, Susan Griffin et Victor Zarate ainsi que John Brazier pour ses conseils sur l'utilisation du SF6D.

Je tiens aussi à adresser quelques mercis personnalisés à Florence Jusot, Marc Perronnin et Paul Dourgnon pour nos collaborations scientifiques ainsi que pour leurs précieux commentaires sur des précédentes versions de cette thèse ; à Juan Pedro Rosa Dias pour nos discussions sur la mesure des inégalités ; à Marion Devaux pour nos nombreux échanges techniques et économétriques ; à Nelly Noël et Aude Sirvain pour leurs talents face aux graphes récalcitrants ; à Hélène Huber pour son aide en ligne face aux secrets de LaTex; à Stéphanie Guillaume pour sa rapidité à transmettre des bases de données ; à Gregory Delsol pour la mission Illustrator ; à Bérengère Davin pour nos échanges réguliers de doctorantes marseillaises en économie de la santé et enfin, Emma Stevens, Stuart Corden et de tout coeur à Lidia Mac Carthy, pour leurs relectures attentives de toutes ces pages écrites dans une langue qui n'est pas la mienne.

Parce que je puise auprès d'eux, confiance, patience et persévérance, je tiens à redire à mes parents, à mes soeurs, Adeline et Roxane, et à Fabien, combien ils comptent pour moi. Ils ont porté avec moi ce projet jusqu'à sa concrétisation, leur soutien n'a eu de cesse.

"Toutes les grandes richesses de ce monde ne valent pas un bon ami" disait Voltaire, j'ai la chance de pouvoir terminer cette phrase au pluriel et je tiens donc à remercier pour leur présence à mes cotés et leur réconfort dans les moments de doute, mes amis.

Cette thèse n'aurait pas été réalisable sans les financements reçus de la MiRE, de l'INSERM, du CNRS, de l'IDEP et de l'Irdes.

Contents

List of Figures

List of Tables

Introduction générale

Depuis plus de deux siècles, les relations entre la santé et la richesse font l'objet de recherches tant en France (Villermé, 1830) qu'au niveau européen (Farr, 1839). Néanmoins, ce sont surtout les travaux des trente dernières années qui ont permis d'analyser le phénomène des inégalités sociales en santé[1]. En particulier, le rapport Black (Townsend & Davidson, 1982) constitue l'un des travaux pionniers sur cette question, non seulement pour le Royaume-Uni, mais aussi pour l'Europe entière. Il a remis en cause l'hypothèse de pauvreté absolue, selon laquelle les "pauvres", ou les ouvriers, ou toute autre catégorie dominée, étaient en plus mauvaise santé, en raison de leurs mauvaises conditions matérielles de vie. Il a alors exposé concrètement que les causes de la distribution sociale de la santé, de la mort et de la maladie ne sauraient être réduites aux conditions matérielles.

Au-delà du constat de leur existence, les mécanismes à l'origine des inégalités sociales de santé ont fait l'objet d'analyses explicatives. A la classe sociale déjà mise en cause pour expliquer les différences de mortalité dans les travaux démographiques (Desplanques, 1984; 1993), de nouveaux déterminants de la santé d'origine et de nature très diverses, mesurés au niveau individuel ou collectif, ont été définis et qualifiés par les épidémiologistes sous le terme de déterminants sociaux de la santé (Marmot & Wilkinson, 1999 ; Berkman & Kawachi, 2000 ; Goldberg *et al.*, 2002). Il s'agit notamment du capital social, de la position relative, des conditions de vie dans l'enfance comme à l'âge adulte. En économie, le champ de l'explication des inégalités sociales de santé a notamment été investi grâce à la méthode de décomposition de l'inégalité en facteurs déterminants introduite par Wagstaff *et al.* (2003). Ces travaux mettent en évidence non seulement une étroite relation entre le revenu et la santé mais aussi une influence du milieu social (appréhendé en général par la catégorie socioprofessionnelle et le niveau d'éducation) et de la protection sociale sur le recours aux soins et sur l'état de santé lui-même (Van Doorslaer & Koolman, 2004). Ces approches permettent donc de mieux cerner les déterminants individuels de la formation des inégalités de santé en considérant les conditions actuelles de vie de l'individu.

[1]Nous distinguons les inégalités sociales de santé qui s'intéressent aux disparités de distribution de la santé au sein de la population des inégalités sociales de recours aux soins qui se concentrent sur les disparités de consommations.

En France, les travaux de recherche portant sur les inégalités dites “sociales” de santé ont été plus tardifs que dans les pays anglo-saxons et scandinaves, alors que, dès les années 70, des recherches démographiques menées par l'INSEE décrivaient les différences considérables de mortalité selon les catégories sociales (Desplanques, 1976 ; 1984). Ces travaux ont permis de confirmer que la France n'est pas épargnée par les inégalités sociales de santé. D'importantes disparités de mortalité sont observées d'une classe sociale à une autre, d'un niveau de revenu à un autre (Mesrine, 1999 ; 2000 ; Kunst *et al.*, 2000; Leclerc *et al.*, 2000). Précisément, ces inégalités sont parmi les plus élevées d'Europe de l'Ouest, notamment en matière de mortalité prématurée de certaines classes sociales (Jougla *et al.*, 2000). De plus, le risque de décès est fortement corrélé au niveau de revenu et cette relation est vérifiée tout au long de la distribution des revenus, indépendamment de l'effet des catégories socioprofessionnelles (Jusot, 2003). Les inégalités sociales liées à la fréquence des problèmes de santé s'observent aussi pour des pathologies très variées (Cambois & Jusot, 2007), telles que les maladies respiratoires, cardiovasculaires, les affections de longue durée, etc. ainsi que pour les trois niveaux constitutifs du handicap, à savoir déficiences, incapacités et désavantages (Boissonnat & Mormiche, 2007). Parmi les déterminants sociaux de la santé, les travaux français documentent particulièrement bien les effets d'une inégale exposition aux facteurs de risque (Goldberg, 1997), d'un recours inégal aux soins, notamment du point de vue de la prévention ou du dépistage (Lombrail & Pascal, 2005), des trajectoires professionnelles, des conditions et de la pénibilité du travail (Cribier, 1997 ; Barnay, 2005) ou encore, une influence de la catégorie sociale du père sur la santé et le risque de décès du descendant à l'âge adulte (Melchior *et al.*, 2006a).

Il nous semble cependant que l'analyse des inégalités sociales de santé en France appelle, aujourd'hui encore, de nouvelles recherches. La persistance d'inégalités sociales de santé, alors que la France a montré sa politique volontariste en matière de réduction des inégalités de santé en créant la Couverture Maladie Universelle en 2000, démontre que ce sujet n'est pas clos.

Cette thèse ambitionne donc d'apporter un éclairage économique sur les inégalités sociales de santé à partir de données françaises. Elle envisage précisément de compléter les travaux existants à deux niveaux : la mesure de la santé et celle des inégalités sociales de santé.

La mesure de la santé a fait l'objet de recherches dans la littérature internationale qui n'ont pas encore été appliquées à la France. L'approche de la mesure de l'état de santé de manière générale est pertinente dans un contexte d'analyse des inégalités de santé, et fait paradoxalement défaut en France.

Quant aux inégalités de santé, nous envisageons d'élargir à la fois le champ de la méthodologie de mesure des inégalités ainsi que celui des déterminants sociaux de la santé. D'une part, les avancées méthodologiques européennes de la mesure des inégalités sociales de santé restent à mener sur des données françaises. D'autre part, dans le champ des déterminants sociaux de la santé, nous relevons que l'explication de l'état de santé à l'âge adulte selon les conditions de vie dans l'enfance n'a pas fait l'objet d'analyses économiques approfondies sur des données représentatives en population, en France.

L'insuffisance d'une mesure des inégalités appréhendée selon l'état de santé général

La majorité des travaux français sur les inégalités sociales de santé utilise des données de mortalité. Moins nombreux sont ceux qui portent sur la morbidité et les indicateurs globaux de santé. Outre leur exhaustivité, les données de mortalité présentent l'avantage de permettre de suivre l'évolution des inégalités sur de longues périodes rétrospectives et d'effectuer des comparaisons internationales. Cependant, elles n'établissent que le résultat des inégalités sociales et n'éclairent pas les processus sociaux ayant conduit aux disparités de mortalité. De plus, comme le rappellent récemment Chauvin et Lebas (2007), les données de mortalité reflètent peu les changements sociaux les plus récents et sont particulièrement sensibles aux changements intervenus au cours du temps. Les auteurs pensent par exemple à la prévention, au diagnostic et à la prise en charge des maladies. Des données de morbidité ainsi que des enquêtes sur les conditions de vie et les déterminants sociaux de la survenue des maladies sont nécessaires pour progresser dans la compréhension des inégalités sociales de santé. Girard *et al.* (2000) soulignent que tous les indicateurs de santé disponibles dans les enquêtes françaises ne sont pas de qualité égale. Néanmoins, plusieurs travaux confirment l'existence d'inégalités de santé selon la morbidité, qu'elle désigne les maladies, les limitations fonctionnelles ou l'auto-évaluation (Mizrahi An. & Mizrahi Ar., 1997 ; Leclerc *et al.*, 2000).

Au niveau international, les travaux de recherche dans ce domaine ont été marqués par la publication de l'ouvrage de l'Organisation Mondiale de la Santé (WHO, 2002) qui recense les mesures de la santé, qu'elles relèvent du niveau individuel ou collectif. Ces mesures sont des outils essentiels pour la mesure des inégalités de santé ainsi que pour éclairer les décideurs dans les états membres de l'OMS. L'élaboration et la mise en place de politiques publiques s'appuient sur des données chiffrées de santé et les indicateurs de santé sont, dans ce contexte, d'une importance considérable. Bellanger et Jourdain (2004) considèrent justement que l'approche économique de la santé dans sa dimension pragmatique ne peut se concevoir sans recours à des indicateurs de santé pour évaluer les résultats d'un système en fonctionnement, qu'il s'agisse d'indicateurs de santé au niveau d'une population ou d'un individu.

Notre intérêt se porte particulièrement sur les indicateurs qui mesurent la santé individuelle générale. Les nombreuses enquêtes sur la santé qui sont réalisées en France comme en Europe, renseignent généralement sur plusieurs dimensions de la santé individuelle. Cependant, bien que certaines enquêtes génèrent un score numérique représentant synthétiquement la santé individuelle, il n'existe pas d'indicateur général, représentatif de l'état de santé global des individus, qui soit validé et reconnu.

La perception qu'a un individu de sa santé demeure une variable incontournable des enquêtes sur la santé, qui englobe, le cas échéant, toutes les dimensions physiques et psychologiques de la santé. En France, peu d'études appréhendent l'analyse des inégalités sociales de santé à partir de cette variable, les approches épidémiologiques, sociologiques et démographiques rejetant fortement l'aspect subjectif de l'auto-évaluation de la santé. Il nous semble cependant opportun d'utiliser ce type de variable pour mesurer la santé et analyser les inégalités sociales de santé dans une approche économique. Tout d'abord, la santé auto-évaluée est une mesure de la qualité de la vie ayant rapport avec la santé au sens large du terme. Ensuite, les enquêtes longitudinales démontrent son aspect prédictif de la morbidité et de la mortalité. Enfin, des transformations sophistiquées ont récemment été proposées dans la littérature internationale afin de la rendre robuste pour représenter la santé individuelle et de lui fournir de bonnes propriétés pour la rendre utilisable dans les analyses de l'inégalité. C'est le cas par exemple des travaux visant à réduire les biais de reports individuels (Shmueli, 2003) et des travaux de cardinalisation à partir d'une distribution de santé (van Doorslaer & Jones, 2003).

Cette thèse vise donc à élargir le concept de mesure de l'état de santé individuel. Partant de l'expérience française de réticence marquée à l'utilisation de la santé auto-évaluée, nous discutons les qualités et les limites de cette variable, puis nous développons une mesure de la santé à mi-chemin entre une approche objective et une approche subjective de la santé, utilisable pour l'analyse des inégalités. Nous discutons, de plus, les apports des raffinements méthodologiques de la mesure de la santé proposés par la littérature qui n'ont, jusqu'à présent, pas été évalués sur des données françaises.

De nouvelles pistes d'analyse et d'explication des inégalités sociales de santé

L'apport des travaux européens de la mesure de l'inégalité

De nombreux chercheurs étrangers ainsi que les organisations internationales, privilégient depuis quelques années l'étude des inégalités de santé en fonction du revenu. Empruntant pour la plupart des méthodes empiriques sophistiquées, ces travaux permettent non seulement de mesurer les inégalités de santé, mais aussi d'identifier leurs déterminants, et d'évaluer la contribution propre de chaque déterminant à l'augmentation ou à la réduction des inégalités, afin, en dernier lieu, d'orienter la réflexion sur les mesures politiques qui permettraient de les réduire. Traditionnellement, les inégalités sociales de santé en France sont mesurées en comparant les indicateurs de mortalité ou, plus rarement, de morbidité pour différentes catégories socioprofessionnelles. Cependant, la catégorie socioprofessionnelle est un agrégat peu satisfaisant pour l'observateur. En premier lieu, le contenu détaillé d'une catégorie socioprofessionnelle est variable pour un même pays d'une date à une autre. En deuxième lieu, la catégorie socioprofessionnelle rend difficile la classification de la population selon un ordre hiérarchique. Selon Wilkinson (1996), la catégorie socioprofessionnelle est un "artefact", mais ne saurait constituer un déterminant de l'état de santé. L'inégalité de santé par catégories socioprofessionnelles peut provenir d'effets culturels ou d'effets matériels. Dans ce dernier cas, la catégorie socioprofessionnelle ne fait que refléter un effet du revenu.

Pour ces raisons, les économistes préfèrent le plus souvent étudier, quand les données le permettent, l'inégalité de l'état de santé en fonction du revenu. Etudier les inégalités de santé en fonction du revenu ne signifie pas nécessairement qu'on préjuge un lien de causalité entre revenu et santé, mais qu'on mesure les différences de niveau de santé moyen par groupes de revenu.

La causalité directe entre revenu et santé a été démontrée (Dourgnon *et al.*, 2001). Dans les ménages pauvres, la probabilité de vivre dans des conditions mauvaises pour la santé (logement ou travail), ou le renoncement aux soins pour raisons financières, est supérieure à celle des individus des ménages riches.

Mais, outre cette causalité directe entre revenu et santé, d'autres mécanismes peuvent expliquer l'existence d'une disparité de l'état de santé selon la position dans la distribution des revenus. Par exemple, les individus moins éduqués sont à la fois plus pauvres et ont plus de risques d'être en mauvaise santé.

Enfin un autre mécanisme est lié au fait qu'un état de santé plus mauvais peut expliquer un moindre revenu, parce qu'il implique quelquefois de diminuer ou d'arrêter une activité professionnelle et conduit alors à un moindre salaire. Cet effet, dit effet de sélection, est documenté, sur données longitudinales américaines par Smith (1999).

Nous proposons d'analyser l'inégalité de santé en fonction du revenu en tenant compte de ces mécanismes divers, c'est-à-dire en mesurant le rôle respectif des différents facteurs associés au revenu et ayant un impact sur la santé. En France comme en Europe, un consensus s'est dégagé sur l'importance de l'étude des inégalités de santé en fonction du revenu. Ainsi, le rapport Atkinson (Atkinson *et al.*, 2001) sur l'inclusion sociale dans la Communauté Européenne soulignait que la réduction des inégalités d'état de santé en fonction du revenu contribue à améliorer l'inclusion et la cohésion sociales, objectif soutenu par la Communauté. De ce fait, l'intérêt s'est porté sur une démarche d'analyse centrée sur la mesure des inégalités de santé en fonction du revenu, qui prend en compte à la fois les effets démographiques et les effets d'autres facteurs eux-mêmes corrélés avec le revenu, comme la catégorie socioprofessionnelle, l'éducation, ou l'activité, et évalue la part de ces différents facteurs dans l'explication des inégalités constatées.

Dans ce contexte, la méthodologie développée par Wagstaff *et al.* (2003) semble particulièrement prometteuse. Elle permet en effet de décomposer simultanément les inégalités constatées en différents facteurs explicatifs (facteurs démographiques mais aussi éducation, activité, région de résidence, etc.) et de mesurer la contribution relative de chacun de ces facteurs à l'inégalité. La décomposition de cette inégalité par la méthode de la contribution relative des facteurs permet de suggérer les politiques *a priori* les plus efficaces de réduction des inégalités de santé en fonction du revenu.

Cette thèse propose de mener une analyse économique des inégalités sociales de santé en France en s'appuyant sur deux types d'outils.

D'une part, elle emprunte les outils de la dominance stochastique issus des nombreux travaux s'intéressant aux inégalités de revenu. Ils permettent d'enrichir la mesure des inégalités de santé de ces connaissances et de décrire de manière robuste et assez directe l'existence d'inégalités.

D'autre part, à l'instar des recherches menées dans les pays anglo-saxons et scandinaves sur la thématique des inégalités sociales de santé, elle met l'accent sur l'appropriation, l'application et l'adaptation de la méthodologie développée par Wagstaff *et al.* (2003) au contexte français.

Elle contribue à inscrire la France dans ce courant important de la recherche au plan international, soutenu aujourd'hui tant par l'OMS, l'OCDE que par la Communauté Européenne.

L'intérêt croissant pour la compréhension des inégalités sociales de santé dues aux conditions dans l'enfance

Plusieurs pistes de réflexion ont été ouvertes dans les travaux internationaux quant aux caractéristiques qui expliqueraient les mécanismes selon lesquels les inégalités sociales de santé se construisent. Alors que des mécanismes sociaux, professionnels, psychologiques ou comportementaux ont pu être mis en cause sur des données françaises, les effets à long terme des conditions de vie dans l'enfance ont été peu identifiés. Seules trois études épidémiologiques récentes mettent en évidence l'influence de la catégorie sociale du père sur l'état de santé et du risque de décès du descendant à l'âge adulte, à partir d'échantillons de données particuliers, à savoir la cohorte épidémiologique de salariés volontaires d'EDF-GDF, dite cohorte GAZEL (Hyde *et al.*, 2006; Melchior *et al.*, 2006a) et l'enquête Histoire de vie (Melchior *et al.*, 2006b).

Pourtant, la question des relations entre les conditions de vie durant l'enfance, les difficultés d'adaptation sociale et la vulnérabilité face aux problèmes de santé à l'âge adulte a fait l'objet de nombreux travaux, pour la plupart épidémiologiques. Plusieurs enquêtes longitudinales ont confirmé l'effet du statut socioéconomique de la famille sur la santé des enfants, mais aussi sur la santé à long terme des individus (Barker, 1997 ; Wadsworth, 1999 ; Barker *et al.*, 2001 ; Power & Hertzman, 1997). Ces recherches ont, par exemple, permis d'observer la relation entre le poids à la naissance et les risques de décès liés à des maladies cardiovasculaires à l'âge adulte. De même, elles ont montré l'interaction entre les conditions de l'enfance et celles de l'âge adulte, ainsi qu'entre des variables de natures socioéconomique et biologique. Cependant, il n'existe pas de données longitudinales françaises de qualité similaire. De même, jusqu'à présent, aucune enquête en coupes transversales ne dispose à la fois d'indicateurs de santé généraux et de données sur les parents des personnes enquêtées. Les récentes données de l'enquête SHARE 2004/2005 (*Survey on Health, Ageing and Retirement in Europe*) comblent cette lacune et permettent de creuser cette piste de recherche sur l'importance des conditions de vie dans l'enfance en France.

L'analyse de cette détermination sociale est d'autant plus importante que le rapport de la Banque Mondiale de 2006 (World Bank, 2005) souligne l'attention primordiale à accorder aux inégalités des chances. Les effets à long terme sur la santé des conditions de vie témoignent de la précarisation précoce d'une partie des membres de la société et met en évidence des inégalités particulièrement injustes. L'influence sur l'état de santé à l'âge adulte des conditions dans l'enfance constitue des circonstances indépendantes de la responsabilité individuelle (Dworkin, 1981 ; Arneson, 1989 ; Roemer, 1998) et représente de ce fait des inégalités des chances en santé.

La thèse envisage que la corrélation entre le milieu social d'origine et l'état de santé à l'âge adulte pourrait également être expliquée par une caractéristique peu explorée : la longévité des parents. Cette hypothèse formalise l'idée que l'état de santé des ascendants influence l'état de santé des descendants. Elle s'appuie sur de récentes analyses qui confirment l'influence de la santé des parents sur la santé des enfants (Case *et al.*, 2002 ; Llena-Nozal, 2007) tout en proposant une perspective d'analyse originale : la persistance de cet effet de la santé des parents sur la santé des descendants à l'âge adulte.

* *
*

Plan de la thèse
Cette thèse est composée de cinq chapitres qui s'articulent en deux parties.

La première partie s'intéresse à la mesure de la santé, qui représente un préalable nécessaire à l'analyse des inégalités sociales de santé. Deux chapitres développent les aspects de la mesure de la santé.

Le premier chapitre offre une relecture des indicateurs de santé couramment utilisés dans les analyses empiriques. De récents travaux révèlent la pertinence de l'information de santé individuelle contenue dans ces indicateurs. Nous revenons sur ce débat et nous mettons en évidence, de manière étendue, les limites et les avantages de l'utilisation de la mesure de santé auto-évaluée. A la lumière des récentes pratiques dans les questionnaires d'enquêtes individuelles et des raffinements méthodologiques proposés dans la littérature internationale, nous retenons l'intérêt de tendre vers une mesure de la santé dans une perspective globale, complétée d'une correction des biais individuels, qui fait pour le moment défaut dans le contexte français.

Le second chapitre propose une méthodologie originale de mesure de la santé, dans un cadre permettant la prise en compte de différentes dimensions de la santé. A partir des données individuelles de l'Enquête Santé et Protection Sociale de l'IRDES, nous développons une mesure de la santé basée sur la somme du nombre de maladies affectant l'individu, pondérée de leur degré de sévérité. Notre construction s'appuie sur un modèle de régression à variable latente expliquant l'état de santé auto-évalué, tout en contrôlant différents facteurs individuels sociaux et de santé. L'originalité et les qualités de cet indice de santé sont ensuite considérées par comparaison avec les méthodes proposées dans la littérature pour mesurer globalement la santé. En outre, l'indice de santé proposé a l'avantage de s'appuyer sur des variables collectées classiquement dans les enquêtes sur la santé.

Nous disposons à l'issue de cette première partie d'un outil de mesure de la santé d'utilisation aisée pour l'analyse des inégalités sociales de santé, qui présente de bonnes propriétés comme la cardinalité.

La seconde partie est composée des trois derniers chapitres. Alors que le troisième chapitre introduit les instruments de mesure dont les analyses font usage par la suite, les deux derniers chapitres présentent les résultats empiriques de cette recherche. Le chapitre

4 fait usage des données individuelles de l'Enquête Santé et Protection Sociale de l'IRDES, considérée en coupes transversales. Le chapitre 5 s'appuie sur la partie française des données de l'enquête européenne SHARE menée en 2004/2005, dont l'échantillon concerne des individus âgés de 49 ans et plus. En outre, les deux chapitres ont recours à l'Enquête sur la Santé de l'INSEE réalisée en 2003, afin de disposer des scores du SF36.

Le troisième chapitre adopte le cadre d'analyse des inégalités de distribution du revenu, afin d'éclairer la perspective de l'étude des inégalités en santé. Etant donné que les inégalités sociales de santé s'appuient sur les deux caractéristiques individuelles que sont le statut socioéconomique, mesuré par le revenu, et la santé, nous envisageons les différents liens que peuvent entretenir ces deux attributs du bien-être et exposons les critères usuels de dominance correspondants. Alors que ces premiers outils permettent de classer des distributions afin de conclure sans ambiguïté sur l'existence des inégalités, ils nécessitent d'être complétés d'une évaluation de la différence d'inégalité entre les distributions. Dans le cadre des inégalités de santé, cet objectif peut être atteint par l'utilisation d'indices spécifiques comme l'indice de Gini appliqué à la santé ou l'indice de concentration. Nous présentons donc plusieurs de ces outils. En outre, nous discutons de manière détaillée et critique l'indice de concentration sur lequel l'analyse menée au chapitre 4 est basée.

Le quatrième chapitre procède à l'étude des inégalités sociales de santé en France en 2004, puis au cours de la période 1998-2004 à l'aide de l'indice de concentration décomposé. Parallèlement, il propose une utilisation originale d'indicateurs de santé sophistiqués et utilisables dans la mesure des inégalités. Ce chapitre apporte des résultats à deux niveaux.

Le premier niveau relève de l'analyse des inégalités. Cette analyse met en évidence des inégalités sociales d'état de santé à l'avantage des groupes sociaux les plus favorisés. En 2004, ces inégalités de santé sont montrées à partir d'une analyse de dominance stochastique puis d'une approche par indices de concentration décomposés. La méthode de décomposition montre qu'en 2004, les plus fortes contributions à l'existence des inégalités provenaient du niveau de revenu, du niveau d'éducation et de la classe sociale. En effet, on constate que plus le niveau de revenu ou d'éducation (baccalauréat et au-delà) est élevé, meilleur est l'état de santé. Le fait d'être cadre ou technicien est aussi favorable à un meilleur état de santé. Ces variables sociales sont des paramètres d'accroissement des inégalités en faveur des couches supérieures. Ainsi, notre recherche met en évidence le fait que le revenu n'est pas le seul facteur important de l'existence des inégalités et que d'autres facteurs sociaux expliquent fortement ces inégalités. Du point de vue de la comparaison au cours de la période, l'analyse met en évidence que les inégalités sociales de santé ont diminué entre 1998 et 2004. Cette diminution s'explique particulièrement par une plus faible élasticité de la santé avec le revenu et une amélioration du niveau de revenu des individus âgés de 56 à 65 ans, ainsi qu'un moindre effet des catégories socioprofessionnelles sur l'inégalité sociale de santé en 2004. Nous étudions aussi le rôle de la couverture maladie universelle (CMU) mise en place au cours de la période. Nous observons que ce sont les individus qui en ont le plus besoin qui ont recours à la couverture maladie universelle, c'est-à-dire les individus les plus pauvres et les plus malades. Cependant, la période d'analyse est encore trop courte pour observer une diminution des inégalités sociales d'état de santé grâce à cette réforme, comme cela a pu être mis en évidence par ailleurs sur les inégalités sociales de recours aux soins.

Le second niveau concerne l'influence de la mesure de santé considérée. L'analyse de sensitivité met en évidence le fait que la mesure des inégalités sociales de santé soit quantitativement sensible à la mesure de la santé considérée. Deux aspects sont testés : l'impact de la distribution de santé choisie pour la cardinalisation de la santé auto-évaluée et l'impact du nombre initial de catégories de cette variable de santé auto-évaluée. Si la distribution de santé utilisée pour cardinaliser la variable catégorique d'état de santé auto-évaluée est concentrée sur les bons états de santé et différencie peu les états de santé extrêmes (un très bon ou un très mauvais état de santé), alors elle aboutit à une mesure réduite des inégalités. Quant au nombre de catégories sur lequel est décrite la santé auto-évaluée, plus il est élevé et plus la mesure de l'inégalité est faible. En effet, les individus ayant plus de choix dans la façon d'évaluer leur santé se répartissent davantage sur l'échelle de catégories de sorte que les effets des catégories extrêmes s'estompent. En outre, la mesure de la santé proposée au second chapitre est testée empiriquement.

Le cinquième chapitre s'intéresse à l'influence sur l'état de santé à l'âge adulte du milieu social d'origine et de l'état de santé des parents, approché par leur longévité relative par rapport à leur cohorte de naissance. Trois approches sont empruntées pour étudier l'état de santé à l'âge adulte : une approche non paramétrique basée sur des outils de dominance stochastique, une approche économétrique paramétrique et une approche par indices de concentration.

La première approche permet d'obtenir des résultats en terme de dominance de premier ordre, très robustes. Elle montre que les distributions de l'état de santé des personnes nées d'un père appartenant aux catégories de "cadres dirigeants et professions intellectuelles" et de "professions intermédiaires et forces armées" dominent significativement celles des personnes ayant des parents "agriculteurs", "artisans et ouvriers qualifiés" ou encore "ouvriers et employés non qualifiés". De plus, la distribution de l'état de santé des individus dont le père était "employé administratif ou vendeur" domine significativement celle des individus dont le père était "ouvrier ou employé non qualifié". De même, les individus dont la mère était "cadre dirigeante, de profession intellectuelle, scientifique ou intermédiaire " ou "employée administrative, personnel des services, vendeuse ", ont significativement plus de chances en santé que les individus dont la mère appartenait aux autres catégories.

En ce qui concerne la longévité des parents, la comparaison des distributions conditionnelles à la longévité relative du père ne met en évidence aucune différence de distribution. En revanche, l'analyse conditionnelle à la longévité relative de la mère confirme l'hypothèse de transmission de l'état de santé parmi les personnes âgées de 61 à 68 ans seulement.

L'approche paramétrique confirme et affine ces résultats en raisonnant toutes choses égales par ailleurs. Elle montre un effet de la profession de chacun des parents et de leur longévité sur l'état de santé à l'âge adulte de leur descendant direct. Alors que la profession de la mère semble avoir un effet direct sur l'état de santé du descendant, la profession du père, quant à elle, a un impact indirect passant par la détermination de la profession du descendant. Cette seconde approche valide en outre "l'hypothèse de transmission de la santé", puisqu'un effet direct de la longévité relative de chacun des parents sur la santé à l'âge adulte est observé.

Par ailleurs, l'analyse montre que les inégalités sociales de santé provenant des caractéristiques familiales sont plus faibles que les inégalités sociales de santé liées aux conditions actuelles de l'individu. Ainsi, nous montrons que les inégalités des chances en santé sont d'une amplitude plus faible que les inégalités sociales de santé.

La troisième approche propose un usage peu commun des indices de concentration. Après avoir cardinalisé l'état de santé auto-évalué à l'âge adulte, elle génère deux indices de concentration de la distribution de la santé sur la longévité de chacun des parents, puis deux pseudo-indices de concentration sur la distribution de la profession de chacun des parents.

L'approche par indices de concentration, basée sur la longévité, confirme l'importance de corriger la mesure de l'inégalité de l'âge des descendants afin de ne pas décrire d'effets non corrigés. Elle met en évidence une inégalité des chances de santé en faveur des individus dont les parents ont connu une forte longévité. Cette inégalité est plus forte quand la longévité du père est considérée, ce qui était aussi observé dans la seconde approche.

Les indices de concentration basés sur la profession des parents décrivent une inégalité de santé en faveur des individus issus de milieux plus favorisés, qui est plus importante quand la profession du père est considérée. Ces résultats sont comparés à un indice de concentration basé sur la profession du descendant et ils montrent à nouveau que les inégalités sociales de santé sont plus fortes que les inégalités des chances en santé. En outre, la méthode de décomposition de l'indice de concentration permet de mettre en évidence une contribution plus forte des caractéristiques sociales du descendant, quelle que soit la variable sur laquelle les individus sont classés dans la construction de l'indice de concentration.

Cette recherche conclut à l'existence d'inégalités des chances en santé sur des données françaises. L'analyse en trois approches est particulièrement originale et la concordance des résultats de chacune de ces approches accroît la pertinence des conclusions et la puissance statistique de l'étude.

Part I

Measuring health

Chapter 1

Measures of health status

1.1 Introduction

Any empirical studies on health, especially those concerning inequalities, have to rely on a measurement of health. Survey data offer various health indicators all of which have different properties and so choosing which one to use is not straightforward.

Moreover, several recent studies have questioned the pertinence of the individual health information contained in health indicators.

Prior to the analyses that are proposed in this thesis, we intend to examine the measurements of health status which are involved in empirical studies.

This review is particularly relevant because economists tend to focus on the general rather than on the particular aspects of health. However, besides self-assessed health, there are no general health variables in most of the health surveys. Nevertheless self-assessed health is prone to dispute. Opinions have been divided on the use of self-assessed health for many years. Indeed, its validity for measuring health correctly has often been questioned in literature and evaluated by comparison to other health measures. As a result, several recent studies have dealt with individual variability in self-assessments of health and the existence of declaration biases related to individual characteristics.

This chapter is organised as follows: following a definition of what is a good indicator of general health, the first section presents from a content point of view the different health indicators which have been used in literature as proxies of “true” health. The term of “true” health is often used in literature to define the latent health. Then, as self-reported health is a widely used measure for general health, we question the use of this variable and point out its advantages and limitations in literature, as well as its correlations and discrepancies with other health indicators. The last section considers better ways to measure health, which consist of either improving available indicators or providing new indicators.

1.2 Health indicators: content, correlations and discrepancies

We briefly list some health indicators used as proxies of "true" health in literature. The aim of this subsection is to underline the wide range of health indicators used in empirical studies.

1.2.1 What is a good indicator of general health?

While many facets of health can be identified, such as functional abilities, biomedical status or emotional characteristics, the assessment or measurement of individual health must take all of these into account. However, there is no single measure or one-dimensional measurement scale for the health of an individual. At best, public health professionals rely on crude health indicators coming from data collection. The term health indicator refers to a single summary measure, expressed in quantitative or qualitative terms, and which represents a key dimension of health status, health care or other related factors.

As we cannot expect a health indicator to be objectively measured (as a temperature, for example), researchers postulate the existence of a latent health variable. In literature, several variables have been regarded as drawing closely on "true" health status, which is the latent health variable. Therefore, the key to measuring health is to be able to access the relevant health information (Knaüper & Turner, 2003). It is common to think that mortality or life expectancy are measures of health. However, surveys offer many other health indicators at individual level that have also been used as good proxies for "true" health. This is the case for indicators based on questions related to ability to carry out daily living activities, self-reported ailments, as well as height and body mass index, which are relatively easy to obtain, or physician-assessed health status or clinical interviews, which are more costly to obtain. Furthermore, mortality fails in the evaluation of the influence of recent social changes but are sensitive to changes in prevention, diagnosis and undertaking to reimburse medical expenses of diseases (Chauvin & Lebas, 2007).

1.2.2 Various health indicators to measure "true" health

Mortality or survival

Mortality is considered to be a measure of health which is not based on a personal assessment of health (Anderson & Burkhauser, 1985). In concrete terms, it is obtained by death counts and related rates, and is sometimes defined by age, sex or sub-groups of population. Survival in elderly cohorts has been found to be a very good health measure. In the absence of such cohorts, in some longitudinal household surveys the death of respondents can be identified in national death registrations using first and last name, month/day/year of birth, sex, race or social security numbers and this is how a mortality measure is obtained for the sample (Franks *et al.*, 2003). There are, however, no such longitudinal surveys in France.

Mortality has been widely used as a measure of health (Grossman, 1972; Parsons, 1980) because it is a measure available for all respondents and it is also independent of a respondent's reporting biases. However, there is a debate on the validity of mortality as an appropriate health measure (Haveman & Wolfe, 1984). In fact, mortality is not a perfect measure of health as some deaths occur suddenly and independently of health status. Moreover, some health problems affecting ability can induce a particularly bad health status, but the individual life prognosis is not reduced.

Clinically observed morbidity

According to Murray and Chen (1992), morbidity can be observed through four types of indicators: the first category refers to physical and vital signs, directly related to the presence of a particular disease; the second to physiological indicators (such as laboratory exams and radiography); the third to functional tests (for instance, ability to carry out daily activities), and the fourth to clinical diagnosis. All four categories require a professional's diagnosis or examination which is why few surveys contain them.

In a shorter health care professional's assessment, Bartel and Taubman (1979) describe an ideal health indicator constructed from the presence or absence of a doctor's diagnosis of particular diseases. Nevertheless, authors underline some issues in their definition of this "ideal" health measure. Firstly, severity, cures and remissions of the disease are not considered. Secondly, a person can be ill without being so diagnosed. Then, the diagnosis requires a health care utilisation. Finally, a diagnosis can subsequently be proved wrong.

Baker *et al.* (2004) also consider self-reports of chronic diseases as a good proxy of "true" health and rely on medical records of the incidence of thirteen ailments[1] to construct a "true" health status. Therefore, they investigate the validity of self-reported ailments and self-reported global health by comparing with medical records. However, although it is possible to do so with their data, generally few surveys look at physiological or physical performances. There are several reasons for this: insufficient survey time, cost of obtaining information as it involves a health professional, logistical difficulties of interviewing in houses and difficulties of doing it again for subsequent groups. In any case, when these health indicators of medical reports are available they mainly concern developed countries.

[1]These thirteen ailments are: cancer; diabetes; migraines; stroke; asthma; bronchitis, emphysema; sinusitis; arthritis; back problems; ulcers; cataracts; glaucoma; hypertension.

Self-reported morbidity

In literature, due to a lack of an aggregate health measure or clinically-assessed health measures, self-reported morbidity variables are used in analyses. They are individual reports concerning functional characteristics, health-related behaviour and health-related characteristics. For instance, Groot (2000) considers as "true" health status responses to a question on health problems and disabilities in reference to a list[2].

In labour market participation, reported functional limitations have been used as proxy for "true" health (Bound, 1990; Disney *et al.*, 2004). They can be instrumental activities of daily living and disabilities in activities of daily living (respectively IADL and ADL).

As far as we know, few studies consider health-related behaviours and characteristics as a unique health indicator. There is, however, a study on Bangladesh (Kuhn *et al.*, 2004), which proposes to use body mass index as a proxy of "true" health. The BMI is considered as an indication of poor nutritional health status ($BMI < 16$), and is strongly associated with increased mortality for women. However, it is not associated with any other health measures such as self-reported health, activities of daily living, physical disability or self-reported chronic morbidity.

Health care utilisation

As visits to the doctor and hospital data are factual and countable, they have been used as proxies for "true" health status. The impetus for using these indicators is that they would be less influenced by an individual's perception. Indeed, considering the Ontario Health Insurance Plan (OHIP) records, Baker *et al.* (2004) rely on health status measure by numbers of care contacts and associated diagnosis.

1.2.3 Another health indicator: the paradoxical self-assessed health

Self-assessed health is widespread in surveys, widely used in health studies and is sometimes considered as a valid indicator of "true" health (Butler *et al.*, 1987).

Compared to a medically-assessed health status, defined by health care professionals disregarding subjective self-assessments by patients, an individual's self-reported health assessment is the result of a more complex aggregation process. This process is based on his observed morbidity (defined by the number of illnesses, his physical performance, disability, and treatments prescribed), his health expectations and his interactions with

[2]"Do you have any of the health problems or disabilities listed on this card ? (exclude temporary conditions)". List of conditions on the card: problems or disabilities with arms, legs, hand, feet, back, neck, difficulty in seeing; difficulty in hearing; skin conditions/allergies; Chest/breathing problems, asthma, bronchitis; Heart/blood pressure or blood circulation problems; Stomach/liver/kidneys or digestive problems; Diabetes; Anxiety, depression or bad nerves; Alcohol/drug related problems; Epilepsy; Migraine or frequent headaches; Other health problems.

health care professionals, but it is also part of his social, cultural and health knowledge environment. Nevertheless, self-assessed health is always compared to other health variables to confirm its health content. The following part aims to highlight results on this ability to proxy "true" health. As evidence is given that self-reports also contain information about respondents' own characteristics (education, standard of living, interaction with the health system) and beliefs about what good health is, disentangling these elements from the "true" health status definition is not straightforward (Thomas & Frankenberg, 2002). This is the reason why the increasing use of self-assessed health in empirical studies has also given rise to a wide debate concerning individual bias that self-assessed health could suffer from.

The following two sections respectively highlight results in literature which uphold self-assessed health as a good proxy of health status and those which criticise its use.

Why is self-assessed health used as a proxy of "true" health?

In literature, self-assessed health is widely considered by comparison with health indicators in order to emphasise its properties in terms of health content. Although self-assessed health questions were at first introduced because medically-assessed health status measures were costly to obtain, they increased in popularity because they are strongly correlated with other health indicators used to proxy "true" health.

1. **Self-assessed health predicts survival.**

   Self-assessed health is a good predictor for survival. For instance, by comparing two points in time, Burström and Fredlund (2001) observe a good prediction of self-rated health on subsequent mortality among adults. Those who rate their health as poor are found three times as likely to die (Long & Marshall, 1999). Using the NHANES data set[3], Seibt (1998) finds that respondents having many contacts with doctors and regular health check-ups provide more valid health evaluations as assessed by the correlation between self-reported health and length for survival. Moreover, most of the studies support these results irrespective of the socioeconomic conditions (Kaplan *et al.*, 1988; Idler & Benyamini, 1997; Burström & Fredlund, 2001).

2. **Self-assessed health is correlated with reported morbidity and medically-assessed health status.**

   Self-assessed health reflects various reported morbidity conditions. Indeed, Groot (2000) confirms a strong relationship between self-assessed health and chronic health conditions, specifically concerning related disabilities. As for diseases, Baker *et al.*

[3]The National Health and Nutrition Examination Surveys is conducted by the National Centre for Health Statistics, Centres for Disease Control. These surveys are designed to assess the health and nutritional status of adults and children in the United States through interviews and direct physical examinations. They have been carried out six times between 1971 and 2004.

(2004) show that all the diseases considered, with the exception of cataracts, are significantly correlated with self-reporting worse health. According to Verbrugge (1984), the most plausible reasons for self-reporting a worse health status are people's greater awareness of their diseases, due to earlier diagnoses. Mechanic (1986) underlines this idea and argues that there may be personal predispositions in the perception of illness and bad health status: people who are chronically ill or get used to their diseases, could have a better perception of the subsequent vital risk and could approach closely their medically-assessed health status. Functional disability, particularly the degree which affects activities of daily living, is also central in the formation of subjective health in cross-sectional studies, whatever the age (Idler & Kasl, 1995; Idler & Kasl, 1991). A comprehensive study of rheumatoid arthritis emphasises that greater activity restriction is associated with lower self-assessed health and provides supports for the idea that a positive change in disability level and especially psychological well-being would have a positive effect on self-assessed health (Nagyova *et al.*, 2005).

3. **Self-assessed health is correlated with health care utilisation.**

   Health care use as well as health care cost are found positively associated with those who rated their health as poor (Long & Marshall, 1999; Ware, 1986; van Doorslaer *et al.*, 2002). Long and Marshall's analysis goes even further by calculating performance indices based on the ratio of actual-to-expected cost within each category of self-assessed health. These indices then suggest a more aggressive treatment of those who rate their health as poor. Furthermore, an important health care utilisation can both improve and worsen the self-reported health status. Improvement is explained by rapid treatment whereas worsening can be due to a higher number of diseases diagnosed. Nevertheless, the more one uses the health care services and has contact with health care professionals, the more one knows about his health issues and perceives his morbidity, allowing him to approach his physician-assessed health status (Baker *et al.*, 2004).

4. **Self-assessed health is a better predictor of health than other health indicators.**

   Although in literature self-assessed health is questioned on its health content by comparison to other health indicators, it has also been found to be a better indication of "true" health than other health indicators. Indeed, the analyses of Krueger (1957) on the 1953-1955 Baltimore Health Survey reveal large discrepancies between self-reported morbidity and clinical diagnoses. Results show that the variance in causes of morbidity went from 2% for syphilis to 100% for rheumatoid arthritis. Furthermore, Idler and Benyamini (1997) underline that self-rated health contributes more to supplementary health information than other health indicators, even those determining mortality. Similarly, self-assessed health is found to be a very inclusive measure of health reflecting health aspects relevant to survival which are not covered by other health indicators (Mackenbach, 2002). Moreover, self-assessed health extends the information obtainable from morbidity indicators by describing the quality rather than merely the quantity of functional abilities. It gives insights into matters of human concern such as pain, suffering or depression that could not be deduced solely from medically-assessed health or laboratory tests.

5. **Self-assessed health expresses individual preferences.**

   Self-assessed health is comparable to a quality of life indicator as it focuses on people's feelings about their personal circumstances. Life satisfaction refers to individual subjective assessment, such as self-assessed health, which is individually evaluated compared with a normative reference or according to an individual's own aspirations. Moreover, this subjective health variable gives information about individuals regardless whether they seek care or not, and can thus reflect positive aspects of good health. From this point of view, biases inherent to subjective reports do not threaten the validity of the measurement process; health status or quality of life are such as the individual perceives them. This way of considering self-assessed health raises the political question of how to balance needs against an individual's subjective demand.

Why is self-assessed health called into question?

Self-assessed health seems to be a good predictor of morbidity and mortality under the assumption that individuals rely on mortality and morbidity relevant information and ignore irrelevant information in their judgments. Nevertheless, as subjective health does not focus on a specific dimension of health, it encompasses strong emotional dimensions and self-reports can be distorted at various levels.

1. **Self-assessed underestimates or overestimates "true" health according to morbidity conditions.**

   Doctors and individuals have dissimilar perceptions of an individual's health status. Self-assessed health cannot reasonably be strictly equal to medically-assessed health indicators. However, an individual is expected to use only morbidity-relevant information to evaluate his health status. As health conditions are appreciated differently according to the burden of pathology or to the variations in illness perception, self-perceived health status may be far from "true" health. For instance, respondents who have the flu at the time of interview are likely to assess a health status differing from a more valid health judgment such as length of survival (Seibt, 1998). However, it is rather the psychosocial well-being related to the disease than their acute aspect that seems to influence self-reports of health. Indeed, depression has a negative effect on perceived health even if the mortality risk is not higher, which is particularly observed with men (Rodin & MacAvay, 1992). As for the nature of pathologies or limitations, Groot (2000) picked out three health conditions which significantly increase the gap between a poor and a very poor health: difficulties in hearing; skin conditions and allergies; and heart or blood pressure or blood circulation problems. People with one of these three problems are inclined to say that their health is very poor even if these diseases do not increase mortality risk and are widespread amongst the population. The report of self-assessed health is thus influenced by the discomfort in daily life due to health problems. Indeed, self-assessed health is more influenced by the disability related to the chronic disease than by the chronic disease itself. Self-assessed health can also over-estimate "true" health because of chronic diseases or disabilities from birth, which could increase an individual's tolerance of health difficulties. Therefore, an individual who is used to suffering, would assess a higher health better than peers even if they share similar pathological or functional health statuses. Analogous results are obtained in Ghana (Belcher *et al.*, 1976), as people who miss body parts rarely report it as morbidity. They behave as if the loss of a limb is no longer a discomfort.

2. **Self-assessed health suffers from individual's response effects, which are independent of "true" health.**

   The correlation between self-assessed health and any individual variable can be restricted to two elements: firstly, an existing correlation between "true" health and individual characteristics variable and, secondly, an individual's appreciation for his "true" health. International studies (Bound, 1990) have stressed the difficulties encountered in comparing levels of self-assessed health across individuals, who differ in terms of socioeconomic, demographic, pathological or cultural characteristics. Evidence is given in literature on variations between health indicators and self-assessed health related to socioeconomic and cultural characteristics and not necessarily differences in "true" health. A classic example in this respect concerns a famous Australian study (Mathers & Douglas, 1998), which observes that the Aboriginal people describe their health status as being much better than that of the general population, whereas they also experience the highest incidence rates of major health problems and other health indicators. This shows that there are individual characteristics that influence self-assessed health and take it away from "true" health. Various individual characteristics have been underlined in this context but, from one study to another, the influence of these variables on self-assessed health is of different magnitude or even sign.

- Effects of gender

  For instance, beyond the fact that men and women suffer from different diseases, gender influences self-assessed health. Groot (2000) argues that men would look at other men of their age when assessing their health status. As they observe that mortality among men is higher than among women, it would reduce the perception of their own life expectancy and thus reduce their self-assessed health status. Groot (2000) highlights that when men say their health is fair (respectively poor), woman would rather say their health is good (respectively fair) in the 1995 British Household Panel Survey. However, his results conflict with other results on gender effect. Indeed, Benyamini *et al.* (2000) conclude that self-assessed health has a lower accuracy for woman's health. Given health statuses, they observe disparities in perception of health status, specifically due to gender. Their results explain that women's self-assessed health reflects both life-threatening and non-life threatening diseases. According to Moesgaard *et al.* (2002), these gender differences in health perception could be explained by gender differences in expectations for health. These researchers estimate differences in cut points of categorical responses according to gender, and find higher cut points for women than for men in the population of the United States. This result would imply that women are more likely to report worse health than men, given the same levels of disability, which could be explained by higher expectations for health for women.

- Effects of social characteristics

  The impact of social characteristics on reports of health are unclear and change according to the studies. It can induce over-reporting such that higher income groups would report worse health status than lower income groups, whereas observed morbidity, namely disability, declines rapidly with income (Murray *et al.*, 2000; Murray & Chen, 1992). Jürges (2007) also finds that richer respondents are likely to under-estimate their health status in their self-assessments on German data. However, it has also been shown the opposite: at lower income individuals are more likely to self-assess a poor health status at given clinical health status in analyses using Canadian as well as British data (Humphries & van Doorslaer, 2000; Hernandez-Quevedo *et al.*, 2005). Using the third National Health and Nutrition Examination Survey conducted in the United States, Moesgaard *et al.* (2002) also find that higher income categories are less likely to report difficulties than lower income categories, while expectations for good health increase with income. To explain their result, they rely on what they call "a wishful thinking scenario", according to which

> "wealthier people have a belief that they should be in excellent health and therefore use liberal standards for excellence in reporting on their own health".

Finally, reports of health have also been found insensitive to social characteristics. Indeed, if we refer to van Doorslaer and Gerdtham (2003) using a Swedish data set, the relationship between self-assessed health and mortality would not vary with income or education levels whereas it varies with demographics and disease characteristics.

– Other individual characteristics

  Finally, few studies have investigated comparisons of self-assessed health according to various cultural or racial determinants. In this context, we would like to quote a US study from Ferraro & Farmer (1999), which concludes that although African Americans and Caucasian Americans report a similar number of chronic diseases, when asked to evaluate their health status, black people report a significantly worse health level than white people. Researchers call it the prevalence of health pessimism among black people. Conversely, comparing white and non-white people, Moesgaard *et al.* (2002) conclude that, given the same level of mobility, non-white people are more likely to report a better self-assessed health.

3. **Self-assessed health suffers from financial or justification incentives.**

   As for occupational status, inability to work, fear of unemployment or physical activity would play a role in under-estimation of self-assessed health (Fylkesnes & Forde, 1992). Indeed, when these three conditions are considered together, Fylkesnes and Forde (1992) suggest that self-assessed health reflects the overall interpretation of how people handle the "pain in life". Other results emphasise the existence of a justification bias between reports of poor health and retirement. Poor health status measured by self-assessed health in many empirical studies is found endogenous for retirement decisions and labour force participation among men (Blau & Gilleskie, 2001). As eligibility conditions for social security allowances are contingent upon bad health in the Dutch disability insurance system, individuals are inclined to emphasise their health condition for financial motives. Kerkhofs and Lindeboom (1995) show that among respondents receiving a disability allowance, reporting errors are numerous and systematic. Bound (1990) emphasises disparities between results on the link between health and labour decisions according to the health indicator used. Self-reports of poor health are significantly correlated to retirement and suffer from reporting bias as people justify their retirement decision by health. This endogeneity therefore exaggerates the effect of health on occupation status.

4. **Self-assessed health suffers from survey administration methods.**

   The reliability of self-assessed health has also been called into question due to the survey method. In particular, evidence of the unreliability of this variable has been given repeating the subjective health question in a same survey but in two different parts: before and after a set of health related questions. Crossley and Kennedy (2002) report that more than a quarter of individuals have changed their reported-health status between the two answers. Recently, Clark and Vicard (2007) have also highlighted an effect of the position of the self-assessed question in the Survey on Health, Ageing and Retirement in Europe (SHARE). In this context, individuals self-assess on average a better health status after they answer the detailed health questionnaire. The wording of the self-assessed health question can differ from one survey to another. The two most widespread types of wording are the general question on subjective health[4] (GSH) and the context dependent question where the individual is asked to compare his own health status with that of his peers[5] (so-called age-related subjective health, ARSH).

This section emphasises that although self-assessed health is an interesting instrument to measure "true" health (because it contains valid information on "true" health), it also presents some limits. It is therefore important to distinguish between differences in "true" health and differences in response style when using self-assessed health variables.

The following section aims, firstly, to describe solutions that have been proposed in literature to improve self-assessed health, and secondly, to emphasise the usefulness of this adjustment. Lastly, it considers other solutions that have been proposed to measure "true" health.

1.3 Health status measures: can we measure "true" health in a better manner?

1.3.1 Towards a correction of individual response effects

Wagstaff and van Doorslaer (2000) emphasise that health status measures have to be as independent of individual effects as possible in self-reports. This individual response effect has been evidenced in literature with various terms such as "state-dependent reporting error" (Kerkhofs & Lindeboom, 1995), "scale of reference bias" (Groot, 2000), "response category cut-point shift" (Tandon *et al.*, 2002), "reporting heterogeneity" (Etilé

[4]The GSH question is: "*In general, would you say that your health is excellent, very good, good, fair or poor?*". This question has, for instance, been used in the Canadian Community Health Survey (CCHS) as well as in the US National Health Interview Survey (NHIS).

[5]The ARSH question adds a reference to a group of peers: "Compared to people your age, would you say that your health is excellent, good, fair, poor or very poor?". This question has been used in the US Vulnerable Elder Survey (VES-13) and in the British Household Panel Survey (BHPS) in 1995.

& Milcent, 2006; Shmueli, 2003), "reporting bias" (Hernandez-Quevedo *et al.*, 2005) or "reporting heterogeneity bias" (Bago d'Uva *et al.*, 2006). Many recent studies have invested in the adjustment for possible individual effects as described in the previous sections. Different econometric methods have been tested for this adjustment in various contexts: measurement of inequality in health, employment or retirement models and cross-country comparisons. Two adjustment methods master the use of additional health indicators and the use of vignettes.

A correction based on the inclusion of other health indicators

This first adjustment method is called the proxy-based approach by Etilé and Milcent (2006). It consists of testing variation of self-assessed health with other health indicators which are assumed to describe "true" health closely. For instance, self-assessed health has been tested by comparison to mortality (van Doorslaer & Gerdtham, 2003), to long-standing disability (Disney *et al.*, 2004), to a score from the Hopkins Symptom Checklist (Kerkhofs & Lindeboom, 1995) or to the value of a comprehensive health indicator for Canadian data (Lindeboom & van Doorslaer, 2004). Recent studies have invested in the construction of a synthetic health index based on diagnosed physical conditions, depression treatment, BMI, grip strength, walking speed (Jürges, 2007) or on various self-reported clinical health conditions (Etilé & Milcent, 2006). Technically, these variables are introduced in ordered Probit models as explanatory variables to estimate self-assessed health. Then, models allow a correction of the reporting bias by assuming that the reference scale varies with one or several particular characteristics such as income, labour market states, education or country. Using a similar approach, Cutler and Richardson (1997) evaluate individual health utility from self-assessed health and incorporate morbidity into the measure of health by using data on chronic conditions. Nevertheless, these other health indicators, except mortality which is not always available, are also subject to measurement error. Indeed, health conditions such as diseases (because they are self-reported variables) contain some amount of measurement error, particularly under-reporting (Baker *et al.*, 2004). Measurement error from additional health indicators can thus bias estimated coefficients. For instance, the use of clinical health may be biased towards the rich by social disparities in access to care and therefore may not reflect "true" health. In this context, the analysis of variations between self-assessed health with other health indicators might consider differences in reporting heterogeneity in the two health indicators rather than deviation of self-assessed health from a proxy of "true" health. Moreover, Bago d'Uva *et al.* (2006) underline two disadvantages of this method. Firstly, socioeconomic related variation in self-assessed health which is conditional on the other health indicators is stripped down by the method. Secondly, the information of "true" health that could be contained in self-assessed health which conditional on other health indicators is lost. Therefore, the alternative is to use hypothetical vignettes.

A correction based on the use of vignettes

The use of vignettes is a methodological innovation introduced in health surveys by the World Health Organisation to address the issue of cross-population comparability. A vignette is a description of a concrete level of ability on a given domain that respondents are asked to evaluate using the same categorical scale they have for their own answer concerning their health in this domain (WHO, 2000; King *et al.*, 2003). Generally several vignettes are used, the set of vignettes describing differences only along the dimension of interest (e.g., mobility, pain, etc.) in order to provide multiple anchors on a single latent scale[6]. They are then used to anchor specific questions that are also asked of respondents as self-assessments. Besides country, this method can also correct for reporting heterogeneity in self-reported health across demographic and socioeconomic groups (Bago d'Uva *et al.*, 2006). The use of vignettes relies on two assumptions: *response equivalence* and *vignette equivalence*. Firstly, respondents are assumed to rate vignettes in the same way as they rate their own health. Secondly, domain levels represented in each vignette are supposed to be understood identically by all the respondents, i.e. irrespective of their personal characteristics. From these two assumptions, a measure of health free of reporting heterogeneity can then be defined as individual variation in responses to vignettes represents reporting heterogeneity. Technically, the vignettes-based approach relies on the use of hierarchical ordered Probit models (HOPIT). Indeed, responses to vignettes are used to estimate effects of population or individual characteristics on thresholds of health report. Health report is then used to estimate effects of population or individual characteristics on "true" health (Murray *et al.*, 2000). The use of vignettes is promising because vignettes can be introduced at low cost in survey questionnaires.

1.3.2 Why is the reporting correction advised?

Measuring the health state of individuals is important for public health policy (Tandon *et al.*,2002). Firstly, it is required to detect differences between individuals at a single point of time or to observe longitudinal changes within individuals. Secondly, it is helpful to evaluate the need for health care and, in this context, consequently to measure inequalities in health. Thirdly, it permits prediction of medical expense risk and gives relevant information for health plans. Therefore, there is a challenge to achieve a health status

[6]For instance, "*Mary can talk to one person at a time in a quiet room but struggles to follow the conversation when there are more people or when there is background noise*". The question is "*How much difficulty did Mary have in hearing someone talking in a normal voice from across the room?*". The response categories are "*none, mild, moderate, severe, extreme/cannot do*".

measure reliable for these objectives. This reliability, especially when using survey data, relies on adjustments of reporting heterogeneity, which has a direct effect on health variables involved in public health studies. Several recent studies have shown and measured the implication of the reporting heterogeneity on public health policy.

1. **Effects of reporting heterogeneity on the measurement of inequalities in health**

   The effect of reporting heterogeneity on the measurement of inequalities in health has been extensively analysed as self-assessed health is often involved in these researches. Studies emphasise important implications on the measurement of inequality. The magnitude and the sign of reporting heterogeneity in self-assessed health related to income have been analysed in France (Etilé & Milcent, 2006) and are found significant for any self-assessed health category and for any income level. As the utilisation of self-assessed health may bias the measurement of inequality, correcting for heterogeneity is relevant if the health variable is viewed for an inequality analysis.

2. **Effects of reporting heterogeneity can induce inequity**

   As health status is appreciated differently according to individual characteristics, it could be inequitable to use this health status, for example, to define individual needs for care in contexts where individuals self-report a worse health status than their actual one. Indeed, in this context, if results of these analyses were used to target needy population it could give rise to a biased point of view (Tubeuf & Rochaix, 2007).

3. **Effects of reporting heterogeneity on cross-country comparisons**

   If self-assessed health or any self-reported data suffer from reporting heterogeneity, then these health variables are not comparable across populations and they will not imply the same level of ”true” health. Using the vignettes methods, Kapteyn *et al.* (2007) evidence that about half of the differences observed in rates of self-reported work disability between the Netherlands and the United States can be attributed to reporting heterogeneity in responses.

4. **Effects of reporting heterogeneity on comparisons over time**

   As far as we know, no study has conducted the evaluation of effects of reporting heterogeneity on comparisons over time. However, critical implications for comparisons over time can also be envisaged, because cut-points may systematically shift over time due to rising income, education and health norms. In this context, long term trends may be difficult to assess without the correction of reporting heterogeneity.

1.3.3 Another solution: providing global health indicators

Uses of global health indicators are numerous: to act as a yardstick for spatial or temporal comparisons; to provide evidence to introduce health policy interventions and to identify levels and gaps in the health of a population. Health indicators can thus be considered independently according to the policies' purpose. However, health status is a multidimensional element, composed of different aspects of health and considered as a whole. In this way, policymakers can take advantage of other methods for measuring health that have recently been proposed, such as using synthetic health measures or considering several indicators together.

The "multi-attribute" health measures

The multidimensionality of health motivates the construction of synthetic indicators capable of picturing this multidimensional complexity. Initially, quality of life indicators have been improved over the last twenty years for the same reasons. More recently, generic health surveys have been proposed in order to encompass the patient-reported health information in several health dimensions. They consist of the description of multiple dimensions of health and the use of utility weights for each of these dimensions in order to construct "multi-attribute" health measures. Such generic health surveys are used to construct generic health instruments such as the Short Form 36 Health Status Questionnaire (SF36) (Ware & Sherbourne, 1992), the EuroQol (The EuroQol Group, 1990)[7]. In addition, these generic health surveys can also concentrate either on specific health problems, such as asthma (Asthma Quality of Life Questionnaire - AQLQ, Juniper *et al.*, 1993) or depression (Centre for Epidemiological Studies Depression Scale - CES-D, Radloff, 1977), or on specific health dimension, such as functional (Functional Status Questionnaire - FSQ, Cleary & Jette, 2000) or mental (Mental Health Inventory - MHI, Veit & Ware, 1983).

Furthermore, a recent study (Lindeboom & van Doorslaer, 2004) considers the Health Utility Index (HUI) from MacMaster University (Feeny*et al.*, 1996) as a "gold" health indicator.

> "While this measure also relies on self-reporting, one advantage is that respondents are only required to classify themselves on eight attributes. The overall individual health utility score on a scale of 0-1 is derived using weights which are derived from a different valuation survey on a different sample of individuals. As such, it represents a more valid and reliable general health measure than the single self-assessed question".

[7]From methodological aspects, SF36 physical and mental health scores are built from a factor analysis, which considers a set of eight functional health and well-being scores. EuroQol 5D which is a quality of life-years (QALY's) indicator, is constructed with health utility assessment methods that enable to reveal individual preferences. These generic health indicators use a mix of visual analogue scale and standard gamble. Their availability relies on particular questionnaires and algorithms.

Based on a health questionnaire, the HUI aggregates eight attributes, namely vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain, which are weighted according to the general population considered.

Although these generic health measures would offer interesting analysis perspectives, their availability is usually restricted to specific health interview surveys, which have also very limited information on living and socioeconomic characteristics. Therefore, surveys concerning generic health profiles are often not suitable for the analysis of relationships between income and health. However, some recent analyses try to insert advantages from these generic health measures into other more widespread health variables.

Including several health indicators for a global health measure

According to Bound *et al.* (1999), who emphasise the use of various health indicators to obtain global health information, methods for measuring health in the recent literature involve the use of several health indicators in addition to self-assessed health. Even if the reason behind this motivation is also methodological[8], as continuous health indicators are preferred to widespread ordinal categorical ones in some analysis contexts, many recent studies are interested in changing self-assessed health to a more-informed health indicator.

As the Health Utility Index (HUI) has been compared in recent studies to an "objective" and comprehensive health status measure (Lindeboom & van Doorslaer, 2004), the distribution of this health indicator has been used to reinforce self-assessed health in a Canadian survey (van Doorslaer & Jones, 2003). Following the same method, the Flemish EuroQol 5D (Lecluyse & Cleemput, 2006) has also been used to change the Flemish self-assessed health in a continuous health indicator.

1.4 Conclusion

A perfect health measure does not exist as even less subjective health indicators, such as self-reports of chronic conditions or health care records, are subject to measurement error (Baker *et al.*, 2004; van Doorslaer *et al.*, 2004). Nevertheless, promising methods have recently been proposed to correct these individual response effects. Moreover, we showed that self-assessed health and other health indicators improve the measurement of health when considered together. Although critics of synthetic index believe that this approach mixes apples and oranges, proponents argue that finding connections between dimensions is necessary in making real life decisions (McDowell, 2006). Moreover, health indices are commonly used in economic analyses and policy-decision making.

Therefore, there is a challenge to achieve reliable health status measures. Indeed, measuring health is a necessary precondition for any decisions in health policies. In public

[8] These questions will be treated in chapter 2.

health research, such measures are required for different reasons. The primary purpose is to detect differences between individuals at a single point of time or to observe longitudinal changes within individuals. Secondly, these measures are helpful when evaluating the need for health care and are used in this context to measure the existence of inequalities. Thirdly, they enable us to predict medical expense risk and give relevant information for health plans. Many other uses of health status measures could be quoted.

To conclude, this chapter highlights that health measures have different properties in terms of contents and recommends a multidimensional measure of health, which can be tested and corrected from reporting heterogeneity due to various factors. This is the reason why this dissertation has turned towards the measurement of health through a global indicator presented in chapter 2. This health indicator respects the present conclusions. Firstly, it involves several health indicators; secondly, it corrects for individual response effects and finally, can be used for public policy decisions.

Chapter 2

A new measurement of health encompassing several dimensions of health

The analysis described in this chapter is a joined research with Marc Perronnin (IRDES). Primary results of this study have been published in Issues in Health Economics Series.[1]

2.1 Introduction

The major challenge in measuring health is that the concept of interest cannot be measured directly in its globality; it can only be measured indirectly by indicators such as surveys, or partially, by clinical observations. These indicators are incomplete capturing only parts of the concept to be measured, and sometimes require to be aggregated. The measurement of an individual's health status that approximates his "true" health status is not only a crucial issue, but also one of the most interesting challenges for studies of health economics. Indeed, there are few measures of health which approach health status as a global concept whereas there is an interest to do so (Chauvin & Lebas, 2007).

We rely on the main conclusions of chapter 1 to propose an alternative approach to the measurement of health, which is as close as possible to "true" health. To do so, we will look for a measurement of health halfway between subjective health and less subjective health.

Firstly, we believe that self-assessed health is an interesting element to take into account for this measurement of health. In addition to the advantages underlined in chapter 1, it appears to us that there is also an argument for the use of self-assessed

[1] Perronnin M., Rochaix L., and Tubeuf S. (2006) Construction d'un indicateur d'état de santé agrégeant risque et incapacité, Questions d'économie de la santé n^{o}107. Série Méthodes. IRDES

health to represent individual health that has been ignored until now. According to the philosophers Merleau-Ponty and Bergson, the perception is the ground in which knowledge takes roots[2](Fressin, 1967). Bergson (1920) upholds that we would indisputably know what we are[3] and similarly, Merleau-Ponty (1948) states that the perception is a fact whose evidence is self-sufficient[4]. They do not mean that the perception is exactly identical to "true" health but argue that we are never fully mistaken[5] but can be part of the illusion.

Nevertheless, the will to give a quantitative value to a perception goes against the perception, which has a qualitative knowledge in essence (Fressin, 1967, p.280). Therefore, we cannot rely on self-perceived health only; we aim to construct a quantitative measure of health.

Secondly, we acknowledge that health is hardly objectively measured. The adjective "objective"[6] simply means

> "(i) not influenced by personal feelings, interpretations or prejudice; based on facts; unbiased: an objective opinion,
> (ii) intent upon or dealing with things external to the mind rather than with thoughts or feelings, as a person or a book,
> (iii) being the object of perception or thought; belonging to the object of thought rather than to the thinking subject (opposed to subjective),
> (iv) of or pertaining to something that can be known, or to something that is an object or a part of an object; existing independent of thought or an observer as part of reality."

Finally, the definition of an objective general health measure consists of looking for a health indicator uninfluenced by emotions or personal prejudices, which is based on observable phenomena or factual information. Individual's health status is partly unknown to the individual himself as well as to the health care professionals. Even a doctor himself is unable to evaluate perfectly his global health.

In this context, we define an appropriate conceptual framework to measure health. As we cannot decide which one between the individual and the doctor has the best ability to measure health, we propose to construct a concrete measure of health using both qualitative and quantitative variables from health surveys. In doing so, we suggest the construction of an indicator describing "true" health. Our construction relies on three

[2]" *Bergson et Merleau-Ponty font de notre perception le sol dans lequel la Connaissance plonge ses racines*", Fressin (1967) in la perception chez Bergson et chez Merleau-Ponty, p.277.

[3]"*L'existence dont nous sommes le plus assurés et que nous connaissons le mieux est incontestablement la nôtre, car de tous les autres objets nous avons des notions qu'on pourra juger extérieures et superficielles, tandis que nous percevons nous-mêmes intérieurement, profondément*", Bergson (1920) in l'évolution créatrice, p.1.

[4]"*La perception est un fait dont l'évidence se suffit à soi-même*", Merleau-Ponty (1948), Sens et non-sens.

[5]"*Nous ne nous trompons jamais complètement de bonne foi ; mais avoir un corps et des sens, c'est pouvoir se faire complice de l'illusion, c'est percevoir par une blessure béante où vient parfois s'abîmer la perception, pour échouer dans l'ambiguïté*", Merleau-Ponty (1945), in la phénoménologie de la perception, pp. 303-309.

[6]From (Dictionary.com, 2006)

elements: (i) we assume that the number of diseases and their severity characteristics is the least subjective health information available in surveys; (ii) we assume that the subjective health status contains implicit general health information, and (iii) we control for individual characteristics within a latent variable model.

Our motivation to construct a new health measure also relies on two elements. Firstly, a continuous and cardinal health indicator is lacking in France. Secondly, we have at our disposal a rich health survey containing information on health. The second section presents these elements. The third section describes the modeling strategy of the index of health. The fourth section presents empirical results. Several methods have been proposed in literature to change self-assessed health into a continuum. Generally, these methods impose a scaling assumption on the ordinal categorical variable, which contrasts with our construction. In the fifth section, we compare our methodology with these approaches. Conclusions are described in the last section.

2.2 Aggregating several dimensions of health to measure a general and cardinal health status

Two approaches are followed to obtain a measure of health on a unique scale from multiple indicators. The first approach relies on multidimensional analysis techniques and consists of summarising information provided by different indicators into few factors or into a unique one. This method implicitly assumes that the different dimensions of health are influenced by a common latent variable or are interacting. It is thus advised when all the indicators considered are highly correlated and it consists of a common factor analysis. However, when the indicators are relatively independent this approach induces a reduction of information. As a consequence, the second approach that relies on the aggregation of different dimensions of health might be preferred. Aggregate measures of health are generally based on assessment of individual's utilities with regard to a set of health characteristics. There are various methods used to evaluate these utilities, such as expert rating, individual self-rating, standard gamble or time-trade-off. Unfortunately, these methods heavily rely on specific questionnaires and so, are difficult to implement on a large scale with any dataset.

We believe that a discrete health indicator or an ordinal indicator restricts empirical uses and measuring health on a continuum is preferred.

2.2.1 Why is the continuous aspect desirable?

Three arguments support the continuous aspect.

1. **The preference for numerical indicators**

Our first argument relies on the preference for numerical indicators in empirical reasoning. Numerical health indexes are generally intended for economic analyses of outputs and for comparing results. Indeed, such indicators enable us to calculate synthetic statistics such as means or variance and to construct confidence intervals. They also permit the calculation of a health stock in the population, the graphical representation of detailed distribution or, the decomposition of indices such as concentration indices. Therefore, they permit to draw a distribution analysis.

2. **The limits of dichotomisation**

   Our second argument concerns continuity as opposed to dichotomisation. Any categorical variable can be transformed into a numerical dichotomous indicator by dividing items into two categories. Although this type of indicators is easy to interpret, it provides weak information: an individual is either ill or not. There is thus no gradation in his health status and we cannot describe the distribution of health status as asymmetric, heavy tailed, etc. The dichotomisation clearly induces a loss of information for an initial indicator described in more than two categories. Moreover, the choice of the cut-off point is not straightforward and will influence subsequent use of the health indicator. Considering self-assessed health, Wagstaff and van Doorslaer (1994) have pointed out that the lower the cut-off point, the greater is the degree of inequality.

3. **The need to take into account within-categories differences**

   Our last argument concerns health utility differences within categories of self-assessed health. Indeed, when an individual reports a good health status equal to the category "good health", it does not mean that his health status is strictly equal to the health status of all the other respondents in the same category. Therefore there is a need for a distinction of individual health statuses within categories of self-assessed health. Ideally, this distinction would be done if individual health statuses were defined on a continuum of health statuses.

Finally, we support that the technical foundation of health measurement relies on the ability to rank each individual's health status on a continuous scale.

2.2.2 How can we measure a continuous health variable?

We use data from the 2002 Health and Health Insurance Survey from IRDES (so-called *Enquête Santé, Soins et Protection Sociale*) to get an indicator measuring health on a continuum of health states in France. Considering the abundance of health information contained in this dataset, it is appropriate to rely on it in order to construct a cardinal and general health index. Run annually from 1988 to 1998 and every other year then, the IRDES-HHIS represents data on French households (except those living in overseas territory or those living in "collective housing" such as long-term care hospitals, religious communities and elderly people's homes) and covers about 20,000 individuals in 7,338 households. The IRDES-HHIS provides information on socioeconomic and demographic characteristics as well as on health status and health insurance coverage. Moreover, each household keeps a medical consumption record for one month by filling out a form. All pharmaceutical expenditures, hospital and ambulatory care consultations are also reported. A basic issue in constructing a health measure is how to choose among the large number of information that could potentially be included. We consider two types of information: medical and functional health, and subjective health.

Reported diseases count and severity induced

Diseases are a morbidity indicator that give an important information on health status. In the literature, the self-reported incidence of some ailments have already been used as less subjective than self-assessed health (Baker *et al.*, 2004). In our context, we consider that the individual number of diseases can reinforce information on health coming from self-assessed health.

We exploit the fact that a stock of diseases represents a cardinal indicator. The IRDES-HHIS diseases report depends on a combination of answers to the question "*Which diseases, health difficulties or disabilities do you have at the present time?*" together with a list of disorders provided as a prompter[7]. Thus, a continuous health variable can be constructed from this dataset by summing the total number of diseases per individual. A medical team in IRDES validates the reported morbidity file by considering it as a whole and corrects glaring errors in reports.

Although a sum of pathologies would give interesting information on an individual's health status, a simple sum has important limits. Indeed, it would come to a conclusion that someone suffering from any two diseases is in worse health than someone suffering from any one disease. However, if the second individual is a terminal cancer patient and the first one has for example, diabetes and eczema, it seems essential to balance this sum of diseases. Similarly, a disease sometimes is just an event occurring in life with complete recovery afterwards; whereas it can also become a chronic part of life sometimes resulting in death. It is therefore important to incorporate a severity level to diseases. A good health indicator has to ignore illnesses with very-short term effects. In this context, we choose to measure the extent of physical limitations as well as prevalence of life risk to evaluate morbidity. We identify diseases that individuals have and evaluate the effects of these diseases on quality of life.

[7] The prompter permits limiting the under-declaration of diseases. It is an interesting detail to mention as reports of diseases have been shown biased by social characteristics.

The IRDES-HHIS has the particularity to contain a clinical assessment of each individual file through two health indicators, namely vital risk and disability levels (Mizrahi & Mizrahi, 1985). Each of the reported health data such as diseases, daily treatment, smoking, previous surgery operations, pregnancies etc. except the self-assessed health, are considered by a doctor in order to attribute to each individual a vital risk and a disability level.

The vital risk is a prognosis on life-expectancy for the respondent at the time the codification is done, this morbidity indicator would translate a quantitative aspect of life. However, the disability level represents a degree of difficulties in daily-life activities.

This second health indicator translates a qualitative aspect of life. These individual-level indicators are ordered categorical variables. The vital risk is composed of seven categories whereas the disability level is divided into eight categories. It is assumed that other diseases from which individuals may suffer could only increase the vital risk or the disability level, but in no case to reduce them. The table 2.1 presents these two morbidity indicators.

Vital risk	**Disability level**
0 No vital risk	0 No discomfort
1 Prognosis very weakly pejorative	1 Very weakly hampered
2 Prognosis weakly pejorative	2 Moderately hampered
3 Possible risk on vital conditions	3 Hampered but normal life
4 Prognosis probably bad	4 Limited professional/domestic activity
5 Prognosis certainly bad	5 Highly hampered
6 Undetermined or deceased during the survey	6 No autonomy for domestic activities
	7 Confinement to bed
	8 Undetermined or deceased during the survey

Table 2.1: Two morbidity indicators in IRDES-HHIS: vital risk and disability level (IRDES, *Enquête Santé et Protection Sociale.*)

In order to channel doctors' assessments and to avoid large disparities in the way they assess an individual's vital risk and disability level, minima levels have been developed. Researchers from the IRDES have developed successive tests methods, in close cooperation with doctors and statisticians to generate minimal vital risk and minimal disability level for diseases (Com-Ruelle *et al.*, 1997). They have assigned a minimal vital risk and a minimal disability level to each reported disease in reference to the International Classification of the Diseases ICD-10 and without any other information. These minima levels are thus created prior to the attribution of vital risk and disability level at individual level and intervene at the end of the doctor's assessment process. If the level of one of the two indicators is lower than the minima levels of the most serious disease reported, the doctor is informed of the anomaly on the screen during the data capture. He is then free to modify the levels he has affected.

Each disease is thus positioned on a scale of six minima vital risk graduations (MVR) and a scale of seven minima disability levels (MDL). The table 2.2 describes these graduations. These minima levels provide an indication of a disease's severity feature, as both the minimal vital risk and minimal disability level respectively give information about the decrease in life expectancy and the reduction of activity caused by diseases. We are particularly interested in these minima levels as they allow diseases to be weighted according to severity. We intend to consider diseases listed in the International Classification of Diseases, whose minimal disability level and minimal vital risks have been evaluated by the IRDES researchers.

Minimal vital risk (MVR)	**Minimal disability level (MDL)**
0 No vital risk	0 No discomfort
1 Prognosis very weakly pejorative	1 Very weakly hampered
2 Prognosis weakly pejorative	2 Moderately hampered
3 Possible risk on vital conditions	3 Hampered but normal life
4 Prognosis probably bad	4 Limited professional/domestic activity
5 Prognosis certainly bad	5 Highly hampered
	6 No autonomy for domestic activities

Table 2.2: Minimal vital risk and minimal disability level (IRDES, Com-Ruelle *et al.*, 1997.)

A set of 1,281 diseases has been recorded in 2002. Each of these diseases has a minimal disability level comprised between 0 and 6 and a minimal vital risk comprised between 0 and 5.

	MDL=0	MDL=1	MDL=2	MDL=3	MDL=4	MDL=5	Total by row
MVR=0	351 46,2% 90,0%	135 17,8% 76,3%	164 21,6% 56,9%	78 10,3% 39,0%	28 3,7% 16,6%	4 0,5% 7,0%	760 59,3%
MVR=1	1 20,5% 8,5%	33 21,7% 19,8%	61 37,9% 21,2%	20 12,4% 10%	11 6,8% 6,5%	1 0,6% 1,8%	161 12,6%
MVR=2	5 4,62% 1,3%	4 3,6% 2,3%	40 36,4% 13,9%	38 34,6% 19%	19 17,3% 11,2%	4 3,6% 7%	110 8,6%
MVR=3	1 0,6% 0,3%	3 1,9% 1,7%	23 14,7% 8%	60 38,5% 30%	56 35,9% 33,1%	13 8,3% 22,8%	156 12,2%
MVR=4	0 0% 0%	0 0% 0%	0 0% 0%	4 4,3% 2%	55 58,5% 32,5%	35 37,2% 61,4%	94 7,3%
Total by column	390 30,4%	177 13,8%	288 22,5%	200 15,6%	169 13,2%	57 4,4%	1281

Table 2.3: Correlation between minimal vital risk and minimal disability level

As few diseases have a very high minimal vital risk and/or a very high minimal disability level, we propose to collect together the two last categories for both MVR and MDL. Each square of the table 2.3 contains the number of diseases with the minimal disability level and the minimal vital risk considered.

We test the linear association of these two minima levels and the table 2.3 represents the correlation matrix. Percentages represent column and row percentages. For instance, we observe 135 diseases with a minimal disability level of 1 and a nought minimal vital risk, which represents 76.3% of diseases with a minimal disability level of 1 and 17.8% of those with a nought minimal vital risk.

We also perform the most common statistical tests to identify the relationship between these two ordinal qualitative variables (cf. table 2.4).

Statistic	**DF**	**Value**	**Prob**
Chi-Square	20	900,4817	<.0001
Likelihood Ratio Chi-Square	20	812,2337	<.0001
Mantel-Haenszel Chi-Square	1	591,3311	<.0001
Phi Coefficient		0,8384	
Contingency Coefficient		0,6425	
Statistic		**Value**	**ASE**
Gamma		0,7411	0,0183
Kendall's Tau-b		0,5547	0,0171
Pearson Correlation		0,6797	0,0162
Spearman Correlation		0,6318	0,0188

Table 2.4: Summary statistics for minimal vital risk by minimal disability level

The significance of Chi-square test and the high value of the Pearson correlation (almost equal to 0.7) indicate that the two variables are strongly dependant and tend to rank diseases on a similar pattern. The Gamma coefficient is based on the number of concordant and discordant pairs of observations, its value is significantly different from 0. These tests confirm the linear association of the two variables, which can be either increasing or decreasing. Tests also emphasised that an aggregation of the two variables in a unique indicator is worthwhile for two reasons. Firstly, minimal vital risk and minimal disability level are highly correlated so if they are considered individually in the same regression they would induce multicollinearity. Secondly, the dependence relation between the two variables indicates that the two minima levels assemble around the diagonal such that sets are clearly associated. Our choice is thus to construct a synthetic indicator combining the two dimensions. Moreover, the strong correlation of the two dimensions underlines that the simple sum of categories of the two indicators would not have any sense as it would produce the same calculation twice.

Considering the high correlation between the two dimensions, an aggregation in a classification of possibilities is advisable. The most adapted method is a correspondence analysis, which provides results similar to those produced by factor analysis techniques. It is based on correlation evidence between the two dimensions considered. It has been argued that correlation approaches produce results that vary according to the particular sample used in an analysis (McDowell, 2006). Nevertheless, as our sample is a set of reported diseases in a representative population survey, this use seems less reprehensible. Moreover, as regard to the small number of combinations (30) produced by the two crossed variables, it is not particularly useful to carry out a correspondence analysis, whose main objective is to simplify wide tables. In this context, we propose an analogous reading of the previous correlation table, behind a correspondence analysis and correlation evidence. We observe for each minimal vital risk the corresponding minimal disability level; among the diseases with a given level of vital risk, we observe some levels of disability that are overrepresented. On the diagonal, five sets of minimal vital risk and minimal disability level are clearly associated and they combine similar levels of severity in the two dimensions. Assuming that $k = 1, ..., K$ represents the severity class related to a disease, we define the following severity levels:

- $k = 1$ representing the severity class for which both the minimal vital risk (MVR) and the minimal disability level (MDL) equal nought.
- $k = 2$ representing the severity class for which both the minimal vital risk (MVR) and the minimal disability level (MDL) are low.
- $k = 3$ representing the severity class for which both the minimal vital risk (MVR) and the minimal disability level (MDL) are average.
- $k = 4$ representing the severity class for which both the minimal vital risk (MVR) and the minimal disability level (MDL) are high.
- $k = 5$ representing the severity class for which both the minimal vital risk (MVR) and the minimal disability level (MDL) are very high.

We then considered the remaining sets, these are combinations a low level of minimal vital risk and a moderate or high level of minimal disability or vice-versa. Although the method seems to be done at a rough guess, we propose to make cut-out figures combining both correlation and sample size in order to avoid very small classes. With this method, we ensure that singular but interesting sets of minima levels are also emphasised. Indeed, using a programmed data analysis, these sets would have been included in the diagonal. The last four classes are thus

- $k = 6$, the minimal vital risk (MVR) is nought whereas the minimal disability level (MDL) is high.
- $k = 7$, the minimal vital risk (MVR) is average whereas the minimal disability level (MDL) is very low.
- $k = 8$, the minimal vital risk (MVR) is average whereas the minimal disability level (MDL) is high.
- $k = 9$, the minimal vital risk (MVR) is high whereas the minimal disability level (MDL) is low or average.

The table 2.5 gives a representation of the layout of diseases' severity classes.

<table>
<tr><td></td><td>MDL=0</td><td>MDL=1</td><td>MDL=2</td><td>MDL=3</td><td>MDL=4</td><td>MDL=5</td></tr>
<tr><td>MVR=0</td><td colspan="2">k=1</td><td colspan="4">k=6</td></tr>
<tr><td>MVR=1</td><td rowspan="2" colspan="1"></td><td colspan="2">k=2</td><td colspan="3">k=8</td></tr>
<tr><td>MVR=2</td><td>k=7</td><td colspan="2">k=3</td><td colspan="2"></td></tr>
<tr><td>MVR=3</td><td colspan="3" rowspan="2">k=9</td><td colspan="3">k=4</td></tr>
<tr><td>MVR=4</td><td></td><td colspan="2">k=5</td></tr>
</table>

Table 2.5: Definition of nine possible severity levels for a disease

This classification will give a more accurate estimation when included in estimations, because it permits avoiding multicollinearity. This severity index is thus related to diseases. For each individual, his/her diseases are then counted and recorded in one of these nine sets.

Self-assessed health

Self-assessed health indicators offer a good opportunity to capture individual preferences and thus to aggregate a wide set of health information. Each individual is supposed to make his assessment with regard to his global health (Bergson[8], 1920). This variable is therefore likely to account for the main dimensions of health. For example, Liang *et al.* (1991) highlight that chronic diseases have an impact on functional heath and that both chronic diseases and functional status influence self-assessed health. Chapter 1 has underlined that self-assessed health has been extensively studied by comparison to other health variables to come to a conclusion as to its validity. Collected in surveys, this indicator has a discrete form as it is more practical to ask individuals to choose among a set of items.

In the 2002 IRDES-HHIS, self-assessed health is collected using the following question: *"Could you grade your health status from 0 to 10? (with 0 being the lowest health status)"*. This scale is slightly different from most of all the other self-assessed health questions, which are usually similar to the one promoted by the European Office of WHO (2000) and

[8]La mémoire n'est pas une faculté de classer des souvenirs dans un tiroir ou de les inscrire sur un registre. Il n'y a pas de registre, pas de tiroir, il n'y a même pas ici, à proprement parler, une faculté, car une faculté s'exerce, par intermittences, quand elle veut ou quand elle peut, tandis que l'amoncellement du passé sur le passé se poursuit sans trêve. En réalité le passé se conserve de lui-même, automatiquement. Tout entier, sans doute, il nous suit à tout instant : ce que nous avons senti, pensé, voulu depuis notre première enfance est là, penché sur le présent qui va s'y joindre, pressant contre la porte de la conscience qui voudrait le laisser dehors", Bergson (1920), L'évolution créatrice, p.5

consist of categories from *"very good"* to *"very poor"* [9]. In the IRDES-HHIS, respondents have no explicit reference on which they can base their evaluation, such as a comparison with people of their age or a precise time period, so they position their health according to their own scale. The representation of the distribution of self-assessed health (see figure 2.1) shows that a majority of individuals reports a health level higher than 7.

The distribution is highly skewed and this skewness is also manifest in the inter-category distances, much smaller between levels 7 to 10 than between 0 and 6. In view of the small number of respondents with a self-assessed health status between 0 and 4, these five categories are hereafter grouped together into a single category identified as the lowest one.

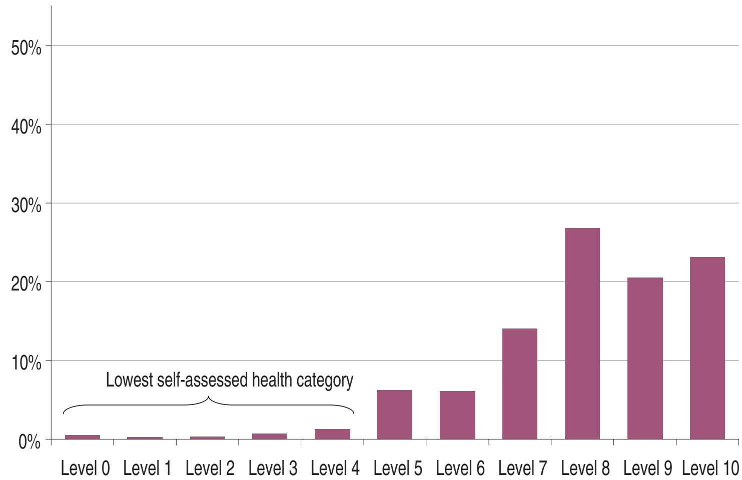

Figure 2.1: Distribution of self-assessed health (*2002 IRDES-HHIS*)

We choose to use self-assessed health as an element of health but we aim to erase as far as possible its disadvantages with the number of diseases.

2.3 A health assessment model

The first step of our analysis is to construct a health measure as close as possible to the "true" health. We believe that the number of diseases combined with the severity levels is a quasi-objective health indicator. We are aware that self-reported diseases can

[9] Considering the distinctive feature of its self-assessed health question, IRDES has recently tried to be comparable with more widespread self-assessed health questions. As a consequence, the 2002 IRDES-HHIS questionnaire introduced a 5-points scale question asked to one half of the sample, along with the usual 11-points scale. A comparison of the two scales has been performed and shows that a score evaluated between 8 and 10 appears to be equivalent to categories *good* and *very good* grouped together (Jusot *et al.*, 2005). Finally in 2004, the 5-points self-assessed health has been asked to all the respondents along with the traditional scale. We have considered this particular feature in chapter 4.

also suffer from individual response judgment. However, the IRDES surveys data have the great advantage to be well-checked by medical experts. In addition, we can also rely on the argument proposed by Jürges (2007), who suggests that diagnosed conditions and measurements are objective health indicators, because diseases are subjective information in factual matters. As a result, we use the number of diseases per severity level to adjust self-assessed health status and so, introduce them as explanatory variables. Our construction relies on Lindeboom and van Doorslaer (2004), who suggest that estimated parameters should be used as weights in their conclusions on the state dependent reporting errors in subjective health measure. Following their suggestion, we investigate an ordered Logit regression explaining the self-assessed health with several individual variables, including the quasi-objective health variables. We then use the estimated parameters to generate the health measure.

In this context, we assume that individuals assess their health considering two issues. Firstly, they score their health with regard to their diseases and the level of severity induced. Secondly, they grade their health by referring to a scale whose graduations are supposed to vary according to their characteristics. In this model, the observed effect of any individual characteristics on self-assessed health is either due to its impact on "true" health or its impact on the responses scales. These two effects cannot be separately identified in the ordered regression model. In order to solve this issue, we assume that the number of diseases combined with severity levels only influences self-assessed health through "true" health and does not influence *ceteris paribus* the responses scale.

2.3.1 The model specification

We shall denote h_{ij}^{subj}, the self-assessed health of the individual i in the household j, and h_{ij}^{*}, the latent variable which represents "true" health according to which the individual i in the household j self-assesses his health. This latent variable is an utility measure, which allows various health dimensions to be aggregated. It is thus a continuous and unobserved variable whereas h_{ij}^{subj} is a discrete dependent variable that takes multinomial ordered values from 4 to 10 [10].

We assume that h_{ij}^{*} is explained by a vector of individual characteristics. Firstly, it depends on D_{ij}^{k} the number of reported diseases of a severity level k, with $k = 1, \ldots, 9$ and $D_{ij} = (D_{ij}^{1}, D_{ij}^{2}, ...D_{ij}^{9})$. We believe that the same illness can have a different impact on the "true" health status. For example, a fractured leg would have more harmful consequences on an elderly person's health status, because of the increased risk of disability induced. Moreover, the older the person, the harder the healing is. Likewise, a same cancer may have different stages of development and cancers from one stage to another are not comparable.

[10] Note that categories from 0 to 4 were grouped in the fourth category.

Therefore, a severity index may not capture the whole "true" health. That is why h^*_{ij} may also depend on X_{ij}, a set of demographic, socioeconomic and health-related behaviour variables, and on an unexplained part. The vector X_{ij} is described in the following subsection. As for the unexplained part, it is composed of two residual terms u_i and ϵ_{ij}, which respectively represent household effects and individual effects taken into account by X_{ij}. This means that the "true" health status of an individual is expressed by the sum of the these two residuals terms and two linear equations, the first one concerning the number of reported-diseases by severity level and the second one containing all the other individual characteristics. This model can formally be written as

$$h^*_{ij} = f_1(D_{ij}, \alpha) + f_2(X_{ij}, \beta) + u_j + \epsilon_{ij} \tag{2.1}$$

On the other hand, we assume that the responses scale of self-assessed health varies with individual characteristics. We denote $c_{a,ij}$, the cut-off points of each category of self-assessed health. The latent health variable h^*_{ij} relies thus on h^{subj}_{ij} as follows.

$$\begin{aligned} h^{subj}_{ij} &= 4 \text{ if } \; -\infty < h^*_{ij} \leq c_{4,ij} \\ h^{subj}_{ij} &= a \text{ if } \; c_{a-1,ij} < h^*_{ij} \leq c_{a,ij} \text{ where } a = 5, \ldots, 9 \\ h^{subj}_{ij} &= 10 \text{ if } \; c_{9,ij} < h^*_{ij} \leq +\infty \end{aligned} \tag{2.2}$$

We assume that the cut-off points $c_{a,ij}$ vary with X_{ij} and with the two residual terms v^a_j and ω^a_{ij} on the adaptative scale g_a. We denote φ_a as a set of coefficients related to each of the covariates in the X-vector, the cut-off points of each category are defined by the following equation.

$$c_{a,ij} \;=\; g_a(X_{ij}; \varphi_a) + v^a_j + \omega^a_{ij} \tag{2.3}$$

In this context, even if individuals have identical levels of "true" health h^*_{ij}, they will assess their health status differently because of their individual characteristics. This can be written as follows.

$$\begin{aligned} h^{subj}_{ij} &= a \text{ if} \\ g_{a-1}(X_{ij}; \varphi_{a-1}) + v^{a-1}_j + \omega^{a-1}_{ij} < \; h^*_{ij} \; &\leq g_a(X_{ij}; \varphi_a) + v^a_j + \omega^a_{ij} \end{aligned} \tag{2.4}$$

If we introduce these assumptions into the expression 2.1, then our model is represented by the following reduced form.

$$h_{ij}^{subj} = a \text{ if}$$
$$g_{a-1}(X_{ij};\varphi_{a-1}) + v_j^{a-1} + \omega_{ij}^{a-1} < \quad f_1(D_{ij},\alpha) + f_2(X_{ij},\beta) + u_j + \epsilon_{ij} \quad \leq g_a(X_{ij};\varphi_a) + v_j^a + \omega_{ij}^a \tag{2.5}$$

Assuming that each previous function is a linear combination of explanatory variables, the equation explaining h_{ij}^{subj} can be written as

$$h_{ij}^{subj} = a \text{ if}$$
$$cst_{a-1} + X_{ij}.\varphi_{a-1} + v_j^{a-1} + \omega_{ij}^{a-1} < D_{ij}.\alpha + X_{ij}.\beta + u_j + \epsilon_{ij} \leq cst_a + X_{ij}.\varphi_a + v_j^a + \omega_{ij}^a \tag{2.6}$$

where cst_{a-1} and cst_a represent constant terms.

We estimate a generalised linear latent model. Our analysis relies on a vector of individual characteristics as well as specific modeling assumptions, which are described in the following subsections.

2.3.2 A set of demographic, socioeconomic and health-related behaviour variables

The model considers some individual characteristics independent of the aggregated health information, namely health-related variables and socioeconomic variables.

Health-related behaviours

Following the point of view of Cutler and Richardson (1997), we assume that health-related behaviours are information of both current and future health status because of their bad effect on health. The incorporation of risk factors is tricky as they are changing overtime and are also by nature included in health indicators. For instance, they interact with chronic as well as mental diseases. Nevertheless, this interaction supports their inclusion in a health model as their effects on health are mediated by other health indicators, such as medical or functional ones (Manderbacka *et al.*, 1999).

Therefore we include in the model, three risk factors, which are available in the dataset: body mass index, tobacco and alcohol consumption[11].

[11] The categories of these three risk factors are constructed behind the questionnaire, they rely on medical assessment (Com-Ruelle *et al.*, 2006; Dauphinot *et al.*, 2006).

Body mass index reflects health status when low as well as when high, and it is associated with elevated risks of mortality and morbidity[12]. Body mass index values can thus be included as a determinant of "true" health. Body mass index is generated with individual height and weight; respondents are classified accordingly, using international references such as underweight ($BMI < 18.5$), normal weight ($18.5 \leq BMI < 25$), overweight ($25 \leq BMI < 30$) and obesity ($BMI \geq 30$). A fifth category is included for missing values.

Tobacco consumption has a long-lasting effect on health related to the quantity and the length of consumption. In IRDES-HHIS, individuals are first asked if they smoke, and if so, they are then asked how many cigarettes they smoke per day, how many years they smoked, whether they smoke at home, whether they are trying to stop smoking and whether they smoked before. Finally, tobacco consumption is divided into four categories: heavy smoker (more than ten cigarettes or five cigars), low (less than ten cigarettes or five cigars), former and non-smoker. A fifth category is introduced for missing values. As for the alcohol consumption, questions are asked on the frequency and the quantity of drinking habits. Another question concerns the frequency with which individuals drink more than six glasses at the same time in a month. In the study, alcohol consumption is also divided into four categories (slight, moderate, heavy and non consumer) and a fifth one for missing values.

Sociodemographic variables

Van Doorslaer and Jones (2003) emphasise the importance to consider a vector of individual characteristics in order to get greater individual-level variations in the health measure. We also believe that individual characteristics when included have a valuable contribution to the control of for reporting bias on reported health.

In addition to health information, the IRDES-HHIS gives detailed social and demographic variables at individual level that we include in our vector of individual characteristics. The table 2.6 describes variables introduced in the analysis.

Concerning demographic variables, 10 age-gender categories are created for men and women aged 16-35, 36-45, 46-65, 65-75 and, lastly 75 and over.

Three levels of education are considered: low (no diploma), medium (primary and secondary education) and high (A-level and more).

The main occupational activity variable has six modalities: employed, unemployed, inactive, homemaker, retired and student.

[12] In order to avoid multicollinearity among regressors, we have excluded obesity and other diseases related to weight from the reported diseases count used to construct the health index. Indeed, these pathologies were not consequences of overweight or obesity on health status but a direct observation of a state of fact. On the contrary, cardiovascular diseases or diabetes are consequences of obesity and overweight so they have been kept in reported-diseases.

Professional activity is also included, namely farmers, craftsmen, executives, technical professions, other employees, skilled workers and unskilled workers. Considering that some individuals (about 17%), do not have an occupational class, for example students and homemaker, because they have never worked, the occupational class of the household head is assigned to them.

In the survey, individuals are asked to report their income in full and/or using an interval scale. When the exact income is missing, the median of the bracket is used. We use the OECD scale[13] to compute the equivalent household income.

Variables	**Mean**	**Proportion**
Age	43.4	
Income (monthly)	1 381.16	
Education level		
Higher education	2,492	28.86%
High school	1,823	21.11%
Secondary education	4,320	50.03%
Professional activity		
Farmer	351	4.06%
Craftsmen retailer	434	5.03%
Executive	1,151	13.33%
Technician	1,926	22.30%
Other employees	2,256	26.13%
Skilled worker	1,667	19.31%
Unskilled worker	850	9.84%
Current activity		
Active	4,986	57.74%
Student	977	11.31%
Unemployed	458	5.30%
Retired	1,541	17.85%
Homemaker	492	5.70%
Inactive	181	2.10%
Social health insurance		
Private	7,766	89.94%
Cmu	291	3.37%
No supplemental insurance	578	6.69%

Table 2.6: Descriptive statistics of sociodemographic variables (*2002 IRDES-HHIS*)

Besides income, education, labour market status and activity status, several health insurance variables are collected, indicating whether the person is covered by private voluntary supplementary health insurance[14] or by a means tested public scheme (Rochaix & Hartmann, 2005). As in 2000, the poorest subgroups of the French population have been granted a limited coverage through the so-called *Couverture Maladie Universelle* (CMU), information also includes whether the individual is covered by a private health insurance beyond compulsory insurance or the CMU complementary insurance in 2002 .

The analysis is also restricted to those in a position to respond to the self-assessed health status question, i.e. those aged 16 and above. Finally, individuals with incom-

[13]The OECD scale gives a weight of 1 to the first adult, 0.5 to the second and subsequent adults and 0.3 to each dependent.

[14]In France, public health insurance is compulsory and universal. It covers about 75% of health expenditures. To finance the remaining part, individual can subscribe a supplementary health insurance, which can be provided through their workplace (being sometimes mandatory) or individually.

plete health questionnaires and those who did not answer some of the sociodemographic questions were also excluded. In the end, the sample contains 8,635 individuals for 2002.

The omitted reference in the analysis is a young man, in employment, highly educated, non-smoker, with a normal weight, who drinks with moderation and has private insurance.

2.3.3 Using individual characteristics to correct the drawbacks of self-assessed health

Chapter 1 emphasises that individuals with the same "true" health status are likely to report different self-assessed health according to their personal characteristics such as age, gender, socioeconomic status and health conditions. We assume that a good health measure should disentangle the "true" health from personal response bias. Therefore, we propose a correction at two levels. The first level is to consider the reporting variation in the thresholds of self-assessed health categories according to individual's characteristics. The second level relies on a random effect, according to which people of the same household are likely to report a similar self-assessed health.

Considering individual variability in self-assessed responses scale

The correction for individual report variability is supposed to allow our indicator to approximate more precisely "true" health. Our testing strategy is in two phases.

Phase 1: Ordered Logit model without varying thresholds
In the first phase, we suppose that the vector of individual characteristics has the same effect on each threshold. In this context, the responses scale are changing through only one translation from one individual and the gap between categories stays the same:

$$\varphi_a = \varphi \tag{2.7}$$

Nevertheless, constant terms still vary with categories a. As a result, we write the following reduced form.

$$\begin{gathered} h_{ij}^{subj} = a \text{ if} \\ cst_{a-1} + X_{ij}.\varphi + v_j^{a-1} + \omega_{ij}^{a-1} < D_{ij}.\alpha + X_{ij}.\beta + u_j + \epsilon_{ij} \leq cst_a + X_{ij}.\varphi + v_j^a + \omega_{ij}^a \end{gathered} \tag{2.8}$$

$$\begin{gathered} \text{i.e. if } \; cst_{a-1} < D_{ij}.\alpha + X_{ij}.(\beta - \varphi) - v_j^{a-1} - \omega_{ij}^{a-1} + u_j + \epsilon_{ij} \\ \text{and } \; D_{ij}.\alpha + X_{ij}.(\beta - \varphi) - v_j^a - \omega_{ij}^a + u_j + \epsilon_{ij} \leq cst_a \end{gathered}$$

It is important to remind that in this model, β and φ cannot be identified and their respective effects on h_{ij}^{subj} cannot be distinguished either. Indeed, the effects of covariates X_{ij} both on h_{ij}^* and on the adaptative scale g_a cannot be separately estimated. Thus, the coefficients may integrate two types of effect, an effect on "true" health and an effect on the responses scale. Moreover we assume that v_j and ω_{ij} are independent of a.

Phase 2: Ordered Logit model with varying thresholds

In the second phase, we allow the thresholds to vary with covariates. Gaps between thresholds are thus supposed to vary from one individual to another. The figure 2.2 explains the reporting process of self-assessed health for two individuals A and B, whose "true" health are represented by respectively H_A^* and H_B^*.

They report their health status according to their own responses scales, which are respectively represented by $C_A^4, ..., C_A^9$ and $C_B^4, ..., C_B^9$. From one individual to the other, the position of the thresholds is varying. This means that each individual positions his "true" health on his own responses scale and reports his health level according to this position. As a result, individual A evaluates his "true" health status h_A^* between C_A^8 and C_A^9 , and reports a self-assessed health equal to 9; whereas individual B evaluates his "true" health status h_B^* between C_B^5 and C_B^6 and reports then a self-assessed health equal to 6. We notice that if individual B had the same responses scale as individual A, he would report a self-assessed health equal to 7.

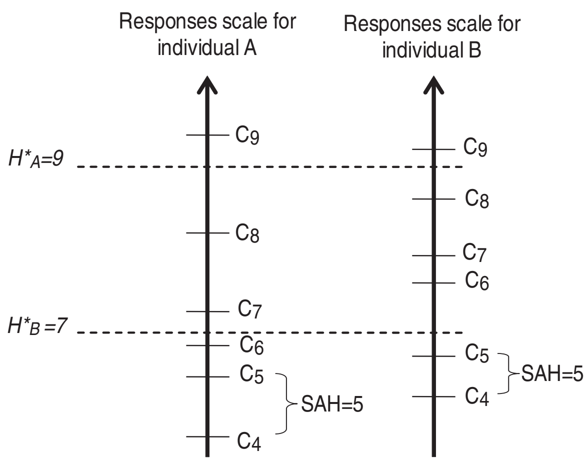

Figure 2.2: Process of self-assessment for health for 2 individuals A and B

We assume a linear specification[15], which allows us to interpret coefficients in the model, easily. Nevertheless, it is important to underline that we cannot distinguish between effects of individual characteristics on "true" health and effects on the scale of self-assessed health. The coefficients integrate two types of effects: an effect on "true" health and an effect on the reporting bias. Moreover, the linear specification does not ensure that thresholds are well-ordered i.e. $g_{a-1}(X_{ij};\varphi_{a-1}) < g_a(X_{ij};\varphi_a)$.

The model gets complicated because we have to write a particular equation for each category, which describes its position according to individual characteristics. In order to avoid significant calculation time, we assume that there is only one covariate that greatly influences the thresholds. We test one by one the effects of each of the covariates on thresholds using an ordered Logit with shifting cut-off points[16]. The likelihood ratio test allows us to select the individual characteristic on which the thresholds vary the most, the lowest log-likelihood. The table 2.7 recapitulates the log-likelihood values of each of these models.

Covariates	**Log likelihood**
Demographic variables	-12,716.96
Education Level	-12,725.35
Occupational activity	-12,700.82
Labor market status	-12,716.76
Household income	-12,735.76
Health insurance	-12,744.16
Smoking	-12,743.63
Alcohol consumption	-12,718.84
Body mass index	-12,740.34

Table 2.7: Effects of covariates on varying thresholds

Among all the covariates, the occupational activity being the variable, which have the highest impact on reporting bias, the log likelihood associated to the model equals $-12,700.837$, whereas it equals $-12,735.764$ for income. In other words, occupational activity is now excluded from explaining variables. We now include cluster effects within the ordered Logit model for health.

Correcting for cluster effect

Unobserved heterogeneity may have several well-known negative consequences on the estimation if it is ignored (Allison, 1999). Indeed, a bias in standard error of estimated parameters leads to an overestimation of the accuracy of statistical test, a lack of effi-

[15] Other specifications are conceivable, for instance an exponential link for differences in thresholds or sequential models, which would be used to estimate $p(SAH \geq k)$ instead of $p(SAH = k)$.

[16] We assume that the introduction of the cluster effect hypothesis in all these regressions is not changing the covariate that greatly influences thresholds. Consequently, we ignore cluster effects in this ordered Logit.

ciency, a heterogeneity shrinkage and a spuriousness bias[17]. We choose to account for this unobserved heterogeneity through a random effect.

Our specification allows to avoid all the previous issues except the spuriousness bias because we use an ordered Logit regression considering random effects[18]. Our motivation to provide for cluster effect relies on the common occurrence in households to report the same self-assessed health for all the members. As shown in figure 2.3, in our sample, more than one quarter of individuals belongs to a household[19] where all the members are reporting the same self-assessed health.

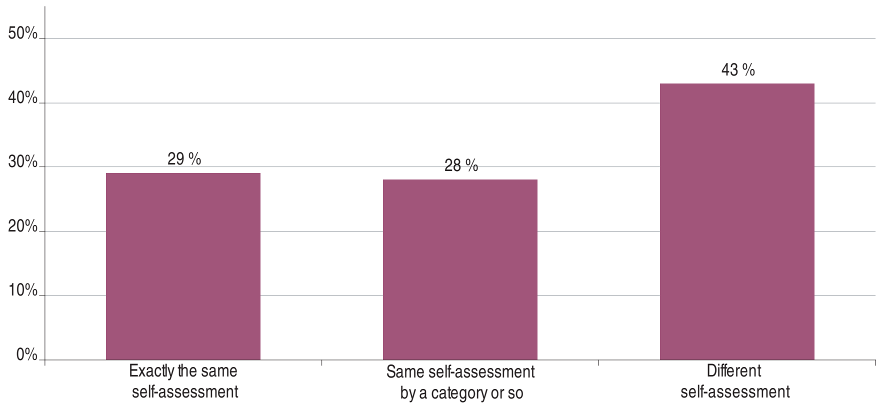

Figure 2.3: Variations in individual self-assessed health within the same household (*2002 IRDES-HHIS*)

As a result, a similar way of reporting health is operated in about 29% of households of more than one individual. It is necessary to highlight that when respondents of same households are not reporting exactly the same self-assessed health status, a quarter of them report a level of health status which differs of one category, only. This cluster effect would be explained either by a similar "true" health status itself, such as genetic endowment, exposition to similar risks for health, similar preferences for health, or similar reporting

[17]The heterogeneity shrinkage means that the variance generated by unobserved heterogeneity attenuates regression coefficients. Spuriousness bias is due to the correlation between household effects and individual effect which bias estimations of the coefficients.

[18]The spuriousness bias could have been corrected by a mixed model, but much more covariates would have been required, leading to unreasonable time calculation. Alternatively, we could have used a fixed effect model to avoid the restrictions on u_i. In particular, unobserved heterogeneity is allowed to be correlated to the covariates, and we thus correct the spuriousness bias due to this correlation. Whereas this type of model is difficult to generalize in non linear cases, an ordered Logit with fixed effect is developed by Ferrer-i-Carbonell and Frijters (2004). Nevertheless, this method presents limitations in our case. Firstly, it discards a considerable proportion of data as it excludes households with no variation in SAH. This exclusion increases standard error since in our sample 30% of individuals are in households with a same SAH level for all the members. Secondly, it does not provide an estimation of variables that are fixed within households, like income by consumption units, which makes our model less informative and more difficult to interpret.

[19]We considered all the households composed of more than one individual.

behaviour, due to cultural factors or similar perception of pain, for instance. Members of a same household are likely to assess their health statuses in a similar way because of common unobservable factors that are not taken into account by the socioeconomic and health variables.

2.3.4 Construction of the health index

The construction relies on the use of the estimated coefficients of each severity level to weight the number of diseases. These coefficients allow us to give a weight, which is not biased by individual responses heterogeneity. The continuous health measure is generated using the combination of diseases by severity level, multiplied by its estimated effect $\hat{\alpha}$ on the latent health variable. For the sake of interpretation, we propose to normalise this continuous health measure in two steps.

In a first step, we choose to normalise each coefficient by $\hat{\alpha}$, which is the estimated coefficient associated to the lowest severity level. The direct use of estimated coefficients $\hat{\alpha}$ as weights would generate arbitrary values as in ordinal regressions, parameters are estimated up to scale[20]. The weight given to a disease of severity level k is thus equal to

$$w_k = \frac{\hat{\alpha}_k}{\hat{\alpha}_1} \tag{2.9}$$

The interpretation of such quantity is straightforward; it represents the number of diseases with the lowest severity level which is needed to produce the same effect on self-assessed than a disease with a severity level k. The health measure can then be written as the sum of all the diseases weighted by the severity level associated with it.

$$I_{ij}^{raw} = \sum_{k=1}^{9} \frac{\hat{\alpha}_k}{\hat{\alpha}_1} D_{ij}^{(Sev_k)} \tag{2.10}$$

This health measure can be compared to a health index as it summarises health into a single number. Our measurement of health combines the medical health and the subjective health controlled by various social dimensions in one instrument. In economic evaluation, these measurements are variously termed "general health status measure" or "measures of health related quality of life". However, we would say that quality of life is broader than our construction. For example, other topics such as daily activities are also considered in the EQ-5D, or such as work and role performance in the SF36.

[20]In particular, their value is sensitive to the distributional assumption for residuals. For example, if we assume that residuals are following a normal law instead of a logistic law, coefficients would be divided by 1,64. In effect, standard normal distribution has a standard error equal to 1 whereas standard Logit distribution has a standard error equal to $\frac{\pi}{\sqrt{3}}$.

In the second step, we change the health measure into a health index described in the interval [0; 1] so as to compare it to other general health status measures such as Health Utility Index or the two summary measures on physical health and mental health from the SF36. In order to do so, we calculate the gap to the highest value it can reach and divide it by the range of its values. This health index can thus be generated using the equation 2.11.

$$I_{ij} = \frac{I_{\max}^{raw} - I_{ij}^{raw}}{I_{\max}^{raw}} \tag{2.11}$$

This health index can be used in its current form in different analyses. Our approach is conservative as we do not include the effects of X_{ij} on h_{ij}^*. Self-assessed health is conditioned on the number of diseases combined with the severity level and we assume that all the socioeconomic variation in self-assessed health is attributed to reporting bias.

Furthermore, we do not account for h_{ij}^* in its entirety. The "true" health status is based in this context both on a medical approach as the number of diseases are taken into account, and on the subjective approach as self-assessed health is considered. Nevertheless, other information could be taken into account to describe all the dimensions of "true" health, for instance, functional characteristics.

The generalised linear latent and mixed model is carried out for equivalent health status, the same diseases and the same severity induced levels.

2.4 Empirical results

2.4.1 Ordered logit models with or without cluster effects

The importance of cluster effects

In a first stage, we estimate an ordered Logit model without variation of thresholds and without cluster effects, and in a second stage we take into account the cluster effect due to the ability to self-assess a similar health status in the same household[21]. The table 2.8 recapitulates the results of these two models.

If we compare results of the two models, we notice that health-related are the parameters whose effects on health are changing the most. For example, overweight and obesity do not have any significant effect on health in the regression with cluster effects whereas these two same variables do have a significant effect on health in the regression model which does not consider cluster effects. A similar pattern is observed for light smokers. Whilst having a high consumption of alcohol has an impact on health in the first model, it does not have such an impact in the second one. The Khi square statistic of the cluster effects parameter equals 210 with one degree of freedom which indicates that inter cluster variance is significantly different from zero. Therefore, it suggests that some unobserved

[21] To do so we use GLLAMMs procedure from Stata software

household characteristics have a strong effect on the global health or on the scale. It is thus relevant to introduce cluster effects in the model as taking into account this unobserved heterogeneity substantially modifies coefficients and their significance. In particular, coefficients associated with the numbers of diseases by severity level are changing. Our decision to take into account cluster effects was motivated by these results.

The following part outlines relevant results concerning the ordered Logit regression with cluster effects and observed for individuals with the same health status.

The impact of health variables on self-assessed health

Regardless of the severity level, for each class of severity self-assessed health is decreasing when the number of diseases increases. The effect on self-assessed health is stronger when the severity level is high.

Being a heavy smoker has a significant and negative impact on self-assessed health. This result is inconsistent with the hypothesis we could have formulated saying that smokers enjoy smoking and increase their well-being by doing so, and that they would self-assess a good health status. It is either that smokers have got bad habits but are conscious of smoking bad consequences on their life expectancy, or they are unconscious that smoking is the cause of their bad health but they suffer from health conditions such as respiratory problems or cardiovascular diseases. For same pathologies, smoking degrades more self-assessed health status.

Not consuming alcohol has a significant and negative impact on self-assessed health. This impact is explained by individuals, who cannot drink alcohol because of medical prescriptions. The fact that individuals do not drink alcohol often stems from a constraint due to health status. Indeed, data do not separate those who do not consume from those who consumed alcohol in the past. Heavy drinkers are likely to report a poor self-assessed health, but this result is not significant.

The impact of the body mass index on the self-assessment of health status is relevant for overweighted and obese people. The higher the BMI, the worse is the self-assessed health. As for smoking habits, individuals who are suffering from overweight must be conscious of the reduction of their ability in daily life, or they suffer from diseases that are consequential to their high weight.

The impact of demographic variables on self-assessed health

Self-assessed health decreases as age increases. Even if results are controlled according to health, the effect of age can be explained by a more pessimistic assessment in older age categories or by an impact of health status which would not be caught entirely. In effect, the same disease can have worse consequences on an elderly person than on a younger person.

Ordered Logit regression without cluster effects					**Ordered Logit regression with cluster effects**				
Log likelihood = -12760,638 Pseudo R2=0,1537					Log likelihood = -12488,537				
Variables	**Coef.**	**S.E**	**P>z**	**[Conf. Int.]**	**Variables**	**Coef.**	**S.E**	**P>z**	**[Conf. Int.]**
Cross product of vital risk by disability									
k=1	-0.338***	0,026	0	[-0,388; -0,287]	k=1	-0,407***	0,032	0	[-0,469; -0,345
k=2	-0.379***	0,045	0	[-0,467; -0,291]	k=2	-0,462***	0,054	0	[-0,568; -0,356
k=3	-0.521***	0,032	0	[-0,585; -0,458]	k=3	-0,668***	0,041	0	[-0,748; -0,588
k=4	-0.792***	0,065	0	[-0,919; -0,666]	k=4	-1,011***	0,077	0	[-1,161; -0,861
k=5	-1.208***	0,144	0	[-1,490; -0,927]	k=5	-1,488***	0,173	0	[-1,827; -1,150
k=6	-0.440***	0,019	0	[-0,478; -0,402]	k=6	-0,539***	0,024	0	[-0,586; -0,491
k=7	-0.327**	0,138	0,018	[-0,598; -0,056]	k=7	-0,301	0,230	0,191	[-0,753; -0,151
k=8	-0.715***	0,102	0	[-0,916; -0,515]	k=8	-0,917***	0,124	0	[-1,160; -0,675
k=9	-0.692***	0,173	0	[-1,032; -0,352]	k=9	-0,953***	0,206	0	[-1,356; -0,550
Tobacco consumption									
No smoker	ref.				No smoker	ref.			
Former smoker	-0,048	0,053	0,365	[-0,151; 0,056]	Former smoker	-0,065	0,065	0,319	[-0,193; 0,063
Light smoker	-0,097	0,063	0,125	[-0,220; 0,027]	Light smoker	-0,187**	0,079	0,018	[-0,342; -0,033
Heavy smoker	-0.392***	0,067	0	[-0,524; -0,260]	Heavy smoker	-0,482***	0,086	0	[-0,651; -0,314
Unknown	0,041	0,081	0,613	[-0,118; 0,200]	Unknown	0,042	0,103	0,685	[-0,159; 0,243
Alcohol consumption									
No cons.	-0.143**	0,055	0,009	[-0,250; -0,035]	No cons.	-0,148**	0,068	0,030	[-0,282; -0,015
Light cons.	ref.				Light cons.	ref.			
Medium cons.	-0,083	0,054	0,125	[-0,189; 0,023]	Medium cons.	-0,045	0,068	0,507	[-0,179; 0,089
Heavy cons.	-0.204**	0,085	0,016	[-0,371; -0,038]	Heavy cons.	-0,172	0,105	0,102	[-0,377; 0,034
Unknown	-0,067	0,094	0,48	[-0,251; 0,118]	Unknown	-0,077	0,116	0,508	[-0,304; 0,151
Body mass index									
Underweight	0,258	0,157	0,102	[-0,051; 0,566]	Underweight	0,198	0,232	0,393	[-0,257; 0,654
Normal	ref.				Normal	ref.			
Overweight	0,122	0,141	0,388	[-0,155; 0,398]	Overweight	-0,234***	0,059	0	[-0,350; -0,119
Obesity	0,100	0,083	0,233	[-0,064; 0,263]	Obesity	-0,575***	0,090	0	[-0,752; -0,398
Unknown	0,011	0,137	0,937	[-0,257; 0,279]	Unknown	-0,026	0,171	0,881	[-0,361; 0,309
Log of inc.	0.163***	0,041	0	[0,082; 0,243]	**Log of inc.**	0,231***	0,061	0	[0,111; 0,350]
Professional activity									
Farmer	-0.311	0,109	0,004	[-0,525; -0,097]	Farmer	-0,423***	0,151	0,005	[-0,718; -0,128
Craftsmen	0.246**	0,101	0,015	[0,047; 0,445]	Craftsmen	0,257*	0,130	0,048	[0,002; 0,512]
Executive	0.276***	0,077	0	[0,125; 0,428]	Executive	0,245**	0,099	0,013	[0,052; 0,439]
Technician	0.143**	0,061	0,02	[0,023; 0,263]	Technician	0,130	0,077	0,092	[-0,021; 0,281
Employees	ref.				Employees	ref.			
Skilled worker	0,091	0,065	0,161	[-0,036; 0,218]	Skilled worker	0,047	0,081	0,564	[-0,113; 0,207
Unskilled worker	-0.208**	0,077	0,007	[-0,358; -0,058]	Unskilled worker	-0,300***	0,096	0,002	[-0,489; -0,112
Education									
Education 3	ref.				Education 3	ref.			
Education 2	0,044	0,059	0,458	[-0,072; 0,159]	Education 2	0,035	0,073	0,636	[-0,109; 0,179
Education less	-0,009	0,060	0,876	[-0,126; 0,108]	Education less	-0,063	0,076	0,403	[-0,212; 0,085
Age crossed with gender									
Male 16-34	ref.				Male 16-34	ref.			
Male 35-44	-0.377***	0,083	0	[-0,540; -0,215]	Male 35-44	-0,617***	0,104	0	[-0,820; -0,413
Male 45-54	-0.879***	0,080	0	[-1,035; -0,723]	Male 45-54	-1,193***	0,099	0	[-1,387; -1,000
Male 55-74	-0.996***	0,145	0	[-1,281; -0,711]	Male 55-74	-1,287***	0,176	0	[-1,633; -0,942
Male=>75	-1.262***	0,176	0	[-1,608; -0,917]	Male=>75	-1,660***	0,219	0	[-2,089; -1,232
Fem. 16-34	-0.206**	0,070	0,003	[-0,343; -0,069]	Fem. 16-34	-0,278***	0,082	0,001	[-0,438; -0,117
Fem. 35-44	-0.371***	0,084	0	[-0,537; -0,206]	Fem. 35-44	-0,600***	0,105	0	[-0,806; -0,395
Fem. 45-54	-0.830***	0,083	0	[-0,993; -0,667]	Fem. 45-54	-1,151***	0,103	0	[-1,353; -0,950
Fem. 55-74	-0.950***	0,141	0	[-1,226; -0,675]	Fem. 55-74	-1,294***	0,174	0	[-1,635; -0,953
Fem.=>75	-1.226***	0,163	0	[-1,546; -0,906]	Fem.=>75	-1,582***	0,202	0	[-1,979; -1,186
Current activity									
Active	ref.				Active	ref.			
Student	0.475***	0,077	0	[0,323; 0,627]	Student	0,570***	0,099	0	[0,377; 0,763]
Unemployed	0,014	0,094	0,878	[-0,170; 0,199]	Unemployed	-0,014	0,114	0,904	[-0,238; 0,21]
Retired	0,027	0,091	0,767	[-0,151; 0,205]	Retired	0,068	0,114	0,551	[-0,155; 0,291
Homemaker	-0,151	0,094	0,108	[-0,335; 0,033]	Homemaker	-0,091	0,111	0,433	[-0,305; 0,131
Inactive	-0.904***	0,151	0	[-1,200; -0,608]	Inactive	-1,068***	0,180	0	[-1 424; -0,719
Social health insurance									
Private	ref.				Private	ref.			
CMU	-0,164	0,124	0,184	[-0,407; 0,078]	CMU	-0,182	0,169	0,281	[-0,513; 0,149
No insurance	-0.336***	0,082	0	[-0,497; -0,175]	No insurance	-0,375***	0,112	0,001	[-0,595; -0,156
Cut-off point estimates									
Cut1	-5,329	0,318			Cut11	-6,532***	0,464	0	[-7,441; -5,622
Cut2	-3,746	0,312			Cut12	-4,659***	0,457	0	[-5,555; -3,764
Cut3	-2,896	0,311			Cut13	-3,617***	0,455	0	[-4,508; -2,725
Cut4	-1,649	0,310			Cut14	-2,053***	0,453	0	[-2,954; -1,166
Cut5	-0,025	0,309			Cut15	0,039	0,451	0,932	[-0,846; 0,923
Cut6	1,193	0,309			Cut16	1,647***	0,452	0	[0,762; 2,533]
Significance of parameters *<0,10, **<0,05, ***<0,01					Intra cluster	1,874	0,1291		

Table 2.8: Results of the ordered Logit regressions without and with clusters effects.

Considering gender, young women assess a significantly worse health status than young men, and inversely in older ages. These results are consistent with previous studies (van Doorslaer & Jones, 2003), particularly those concerning elderly people (Groot, 2000) which were explained in terms of life expectancy. Before self-assessing their health status, men would compare themselves to other men of their age and would observe that mortality among men is higher than among women. Thus, they would give a lower assessment of their own life expectancy and of their health status.

The impact of social variables on self-assessed health

Household equivalent income plays a positive and significant role on self-assessed health; the higher the income level, the better is self-assessed health. Intuitively, as expected, the richest have a better access to the health care system and benefit from a higher quality of cares when they are ill.

Education level has a non-significant impact on self-assessed health whatever the level of education considered.

Concerning the main occupational activity status, being a student has an effect on self-assessed health, which can be compared to the one of age. As age classes are large (16-35 years old), student effect could be explained by a hidden age effect or the absence of particular diseases, such as those due to work conditions. Inactivity, which excludes homemakers, has a negative impact on self-assessed health. That can be explained by both a direct and an indirect health effect. Indeed, in a direct way, individuals out of the labour market at working ages, are likely to be excluded because of their health status. The indirect health effect relies on the fact that an individual in precarious conditions often has a poor health. Finally, unemployment, retirement as well as being homemaker have a non-significant impact on self-assessed health.

Farmers and unskilled workers are likely to assess a worse health status than employees. The common explanation comes from working conditions. Inversely, executives assess a better health status. As we consider individuals having the same health status, an explanation can be found in respect of executives, who may have less health problems because of their higher social status.

Following this idea, having no supplementary health insurance plays a negative role on self-assessed health. That counters to the self selection hypothesis. However, two theories explain this impact on health. Firstly, although people with a lower self-assessed health would have a greater propensity to ask both for care and for supplementary insurance, premiums of this supplementary insurance are more expensive and so, would lead to higher health care expenditures. Secondly, people who cannot afford a supplementary health insurance could be sicker because they cannot have a good access to health care they need, which worsens their health. This first analysis supports the importance of cluster effects. This is why the third model, which considers varying thresholds, includes clusters effects.

2.4.2 Ordered logit model with cluster effects and varying thresholds

As described in the previous method, we choose to make thresholds varying with a unique variable. According to the log-likelihood value of various regression model, occupation status has appeared to be the most relevant. The results of this last model are presented in table 2.9.

These results are similar to those of the previous model with cluster effects but without varying thresholds. However, if we represent the effects of occupational status on the thresholds of self-assessed health, we notice the importance of taking into account varying thresholds.

Figure 2.4 represents the distance from one self-assessed health category to another according to occupational status. It allows us to understand that according to the occupational status, individuals have different levels of health expectations. For instance, the interval of self-assessed health comprised between 9 and 10 is the largest for active individuals, which means that they have a higher probability to self-assess a health status of this level than individuals with other occupational status.

Conversely, retired and unemployed people have lower expectations of good health and are less likely to report a self-assessed health higher than 9. This hypothesis of varying thresholds implies a strict analysis of their effects on health.

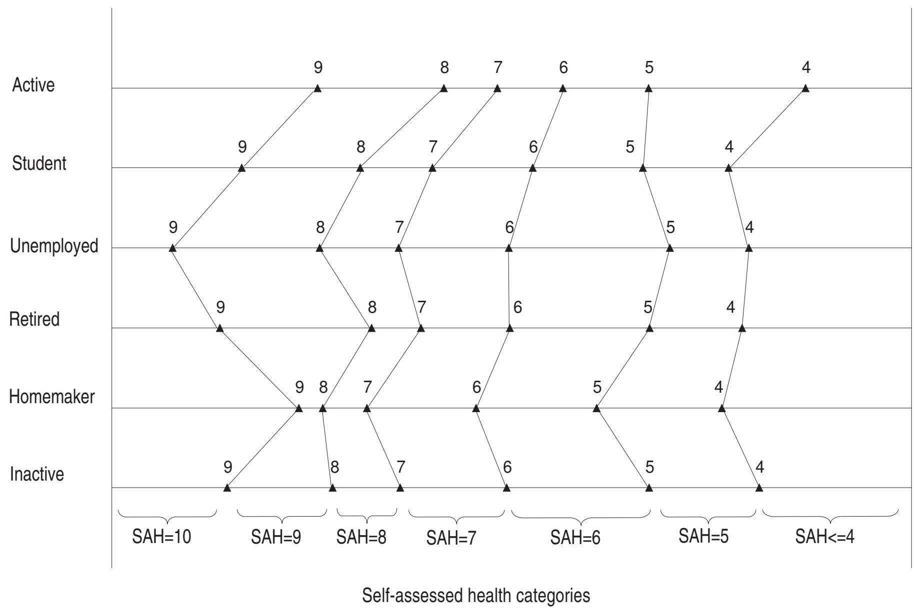

Figure 2.4: Effects of occupational status on the thresholds of self-assessed health.

Ordered Logit regression with cluster effects and occupation varying thresholds

Log likelihood = -12419.813
Condition Number = 269.97443

Variables	**Coef.**	**S.E**	**P>z**	**[Conf. Int.]**	**Variables**	**Coef.**	**S.E**	**P>z**	**[Conf. Int.]**
Cross product of vital risk by disability					**Social health insurance**				
k=1	-0,404***	0,032	0	[-0,467; -0,342]	Private	ref.			
k=2	-0,461***	0,055	0	[-0,568; -0,354]	CMU	-0,175	0,171	0,307	[-0,510; 0,160]
k=3	-0,673***	0,041	0	[-0,753; -0,593]	No supp. ins.	-0,401***	0,113	0	[-0,622; -0,180]
k=4	-1,034***	0,078	0	[-1,186; -0,882]	**Cut-off point estimates**				
k=5	-1,575***	0,175	0	[-1,918; -1,231]	**Cut11**				
k=6	-0,544***	0,024	0	[-0,592; -0,496]	Active	ref.			
k=7	-0,308	0,232	0,185	[-0,763; 0,147]	Student	1,076***	0,409	0,008	[0,275; 1,878]
k=8	-0,946***	0,125	0	[-1,191; -0,702]	Unemployed	-0,111	0,401	0,781	[-0,896; 0,674]
k=9	-0,970***	0,207	0	[-1 376; -0,564]	Retired	-0,817***	0,220	0	[-1,249; -0,385]
Tobacco consumption					Homemaker	0,221	0,309	0,474	[-0,384; 0,827]
No smoker	ref.				Inactive	1,353***	0,313	0	[0,740; 1,966]
Former smoker	-0,069	0,066	0,294	[-0,198; 0,060]	Cons	-6,268***	0,482	0	[-7,212; -5,325]
Light smoker	-0,192**	0,079	0,015	[-0,348; -0,037]	**Cut12**				
Heavy smoker	-0,474***	0,086	0	[-0,644; -0,305]	Active	ref.			
Unknown	0,037	0,104	0,722	[-0,166; 0,240]	Student	-0,145	0,344	0,673	[-0,819; 0,529]
Alcohol consumption					Unemployed	0,585***	0,220	0,008	[0,155; 1,016]
No cons.	-0,146***	0,069	0,033	[-0,281; -0,012]	Retired	-0,195	0,157	0,214	[-0,503; 0,113]
Light cons.	ref.				Homemaker	0,416**	0,203	0,040	[0,018; 0,814]
Medium cons.	-0,050	0,069	0,469	[-0,185; 0,085]	Inactive	1,667***	0,250	0	[1,177; 2,157]
Heavy cons.	-0,170	0,106	0,108	[-0,378; 0,038]	Cons	-4,686***	0,467	0	[-5,601; -3,770]
Unknown	-0,086	0,117	0,463	[-0,315; 0,143]	**Cut13**				
Body mass index					Active	ref.			
Underweight	0,210	0,234	0,369	[-0,249; 0,669]	Student	-0,503*	0,269	0,062	[-1,031; 0,025]
Normal weight	ref.				Unemployed	0,306	0,188	0,103	[-0,061; 0,674]
Overweight	-0,232***	0,059	0	[-0,348; -0,117]	Retired	-0,023	0,140	0,867	[-0,298; 0,251]
Obesity	-0,574***	0,091	0	[-0,752; -0,396]	Homemaker	0,481***	0,169	0,004	[0,151; 0,812]
Unknown	-0,021	0,172	0,901	[-0,359; 0,316]	Inactive	1,456***	0,239	0	[0,987; 1 925]
Log of income					Cons	-3,670***	0,463	0	[-4,578; -2,762]
Log of income	0,233***	0,062	0	[0,112; 0,354]	**Cut14**				
Professional activity					Active	ref.			
Farmer	-0,449***	0,154	0,003	[-0,751; -0,148]	Student	-0,461***	0,165	0,005	[-0,784; -0,138]
Craftsmen	0,278**	0,131	0,034	[0,021; 0,534]	Unemployed	0,044	0,154	0,776	[-0,258; 0,345]
Executive	0,253**	0,099	0,011	[0,058; 0,448]	Retired	0,033	0,128	0,795	[-0,218; 0,285]
Technician	0,136*	0,078	0,080	[-0,016; 0,288]	Homemaker	0,392***	0,144	0,006	[0,110; 0,674]
Other employees	ref.				Inactive	0,846***	0,240	0	[0,376; 1,316]
Skilled worker	0,047	0,082	0,565	[-0,113; 0,207]	Cons	-2,074***	0,459	0	[-2,974; -1,174]
Unskilled worker	-0,307***	0,097	0,002	[-0,497; -0,118]	**Cut15**				
Education					Active	ref.			
Education 3	ref				Student	-0,794***	0,119	0	[-1,027; -0,561]
Education 2	0,029	0,074	0,690	[-0,115; 0,174]	Unemployed	0,005	0,143	0,973	[-0,276; 0,285]
Education less	-0,063	0,076	0,407	[-0,212; 0,086]	Retired	0,310**	0,138	0,025	[0,039; 0,581]
Age crossed with gender					Homemaker	-0,100	0,141	0,480	[-0,377; 0,177]
Male 16-34	ref				Inactive	-0,008	0,263	0,976	[-0,524; 0,508]
Male 35-44	-0,594***	0,104	0	[-0,798; -0,390]	Cons	0,066	0,457	0,886	[-0,831; 0,962]
Male 45-54	-1,156***	0,099	0	[-1,351; -0,961]	**Cut16**				
Male 55-74	-1,255***	0,179	0	[-1,605; -0,905]	Active	ref.			
Male=>75	-1,694***	0,222	0	[-2,129; -1,258]	Student	-0,564***	0,110	0	[-0,779; -0,348]
Fem. 16-34	-0,288***	0,082	0	[-0,450; -0,127]	Unemployed	-0,261	0,160	0,102	[-0,575; 0,052]
Fem. 35-44	-0,597***	0,105	0	[-0,803; -0,390]	Retired	-0,161	0,163	0,324	[-0,480; 0,159]
Fem. 45-54	-1,108***	0,103	0	[-1,311; -0,905]	Homemaker	-0,459***	0,159	0,004	[-0,771; -0,148]
Fem. 55-74	-1,260***	0,176	0	[-1,605; -0,916]	Inactive	0,689*	0,367	0,060	[-0,030; 1,408]
Fem.=>75	-1,604***	0,205	0	[-2,006; -1,202]	Cons	1,718***	0,458	0	[0,820; 2 616]
Intra cluster variance						1,871	0,129		
Significance of parameters *<0.10, **<0.05, ***<0.01									

Table 2.9: Results of the ordered Logit regression with clusters effects and varying thresholds due to occupation status.

2.4.3 The continuous health indicator

The regression coefficients $\hat{\alpha}$ are used as an unbiased weight to construct the health indicator. In a first step, we normalise each estimated coefficient by the one associated to the lowest severity level. The table 2.10 gives weights that are attributed to each severity level according to the modeling concepts and corresponds to the values of the coefficients normalised to the lowest one.

This table can be analysed as "equivalent number of diseases of the lowest severity level": a disease with a severity level of 5, is equivalent to 3.9 diseases with a severity level of 1 in the model with cluster effects and varying thresholds, respectively 3.66 in the second model and 3.57 in the first one. If we represent the distribution of these severity weights according to the model specification, we observe the same pattern whatever the model. However, by comparison to the simplest model, we can see that the correction for cluster effects as well as the consideration of varying thresholds emphasise weights. When the severity is the highest (i.e $k = 5$), the weight associated is the strongest and the model relies thus on varying thresholds and cluster effects. The severity level estimates are particularly different in the model specification when there is an existent level of vital risk. Indeed, there are light differences between severity levels for which $k = 2, 3, 6, 7, 8$ according to the model specifications. For the other values of k, we confirm previous results according to which the cluster effect influences values of coefficients, even when they are normalised by the coefficient associated to the lowest severity level in order to drop the shrinkage effect.

Disease severity level	Without cluster effect without varying thresholds	With cluster effects without varying thresholds	With cluster effect varying thresholds
k=1	1	1	1
k=2	1,12	1,14	1,14
k=3	1,54	1,64	1,67
k=4	2,34	2,48	2,56
k=5	3,57	3,66	3,90
k=6	1,30	1,32	1,35
k=7	0,97	0,74	0,76
k=8	2,12	2,25	2,34
k=9	2,05	2,34	2,40

Table 2.10: Values of weights according to the model specification

Our hypotheses of cluster effects and varying thresholds are thus directly relevant to the health measure, which will be constructed. They emphasise the weight of diseases' severity levels in the indicator and so, the weight of objective health. The raw continuous health indicator can then be generated using equations 2.11 and estimated coefficients of these diseases severity levels. Nevertheless, which estimated coefficients are preferred within the construction of the health index?

Our model specification in three steps has emphasised the importance of cluster effects.

As for the effect of varying thresholds, even if it exists, its implementation is time-consuming and the choice of the covariate on which it is based, depends on the sample considered.

In this context, we prefer to construct our health measure using estimated coefficients from the ordered Logit with cluster effects and without varying thresholds.

The distribution of the constructed continuous health indicator is represented in the figure 2.5, and is compared to the one of the self-assessed health variable. The health index reports an average health equal to 0.89. Generally speaking the distribution of the indicator is concentrated among good health statuses and is spread among bad health.

This health index is synthetical and allows comparisons between different populations. Its continuous aspect enables us to make a distributional analysis, in particular to calculate standard error or confidence intervals. As an example, we propose to consider differences in health status when it is measured by our index, according to gender and age in classes. We calculate the average health index by age classes. As a recent French national survey from INSEE contains SF36 scores, we propose to compare[22] the health index that we have constructed with the SF36 physical score. Results are presented in the table 2.11.

Figure 2.5: Comparison of the distributions of the health index and self-assessed health

Despite their different mean values, both indicators have same patterns according to age and gender. Mean scores significantly decrease with age, the effect being stronger

[22]In order to facilitate comparisons, the health index is described on the $[0; 100]$ interval.

for older individuals. Average differences between health status increases with age; the difference of average health between two age classes being higher because individuals are old. Whatever the score, confidence intervals show that weakening due to age is significant because intervals do not overlap with each other.

Moreover, whatever is the age class, men are healthier than women. This difference is generally significant with the exception of two age classes (35-44 and 45-64 years old) when health is measured by SF36 physical score, and for individuals aged 65 years old and more when health is measured by the constructed indicator. Considering our indicator, the difference between men and women health status increases with age until 64 years old and decreases beyond, which suggests that health, worsens regardless of the gender.

Constructed continuous health indicator (*source:* 2002 IRDES-HHIS)	**Sample**	**Mean**	**Confidence interval 95%**	**Minimum**	**Maximum**
Whole sample	8,635	89.41	[89.14;89.68]	0	100
Women	4,419	87.23	[86.85;87.61]	40.77	100
16-34 y.o	1,572	93.75	[93.39;94.11]	26.52	100
35-44 y.o	879	91.60	[90.97;92.23]	15.54	100
45-64 y.o	1,309	84.87	[84.17;85.57]	19.94	100
65-74 y.o	399	76.78	[75.29;78.27]	29.6	100
75 and more	260	71.09	[69.67;72.85]	0	100
Men	4,216	91.08	[90.71;91.45]	0	100
16-34 y.o	1,513	95.58	[96.31;96.85]	51.81	100
35-44 y.o	845	94.44	[93.92;94.96]	16.41	100
45-64 y.o	1,282	88.97	[88.33;89.61]	14.69	100
65-74 y.o	378	78.57	[76.90;80.24]	5.52	100
75 and more	198	72.19	[69.67;74.71]	0	100
SF36 physical health score (*source:* 2003 French Health Survey, INSEE)	**Sample**	**Mean**	**Confidence interval 95%**	**Minimum**	**Maximum**
Whole sample	20,574	50.43	[50.11;50.70]	5.85	75.98
Women	10,899	49.95	[49.60;50.30]	5.85	72.9
16-34 y.o	3,248	53.85	[53.57;54.13]	9.82	72.9
35-44 y.o	2,395	52.67	[52.34;53.00]	10.02	70.08
45-64 y.o	3,588	49.21	[48.89;49.53]	12.26	71.64
65-74 y.o	1,039	42.11	[41.46;42.76]	10.99	64.93
75 and more	629	36.70	[35.82;37.58]	5.85	63.02
Men	9,675	50.91	[50.72;51.50]	9.74	75.98
16-34 y.o	2,856	54.64	[54.38;76.32]	12.59	70.83
35-44 y.o	2,112	53.07	[52.34;53.80]	12.51	69.98
45-64 y.o	3,282	49.84	[49.52;50.16]	9.74	75.98
65-74 y.o	956	44.44	[43.80;45.08]	10.79	66.92
75 and more	469	39.25	[38.24;40.26]	10.89	61.58

Table 2.11: Comparison of health status when measured by the health index and when measured by the SF36 physical health score, by gender and age.

Considering that different statistical methods have been proposed to transform the ordinal categorical self-assessed health into a cardinal measure, it is interesting to compare our construction with this literature.

2.5 Comparisons with other constructions

In the literature, three solutions highlight the scope of methods proposed to transform an ordered categorical indicator into a continuous one. They assume that the categorical ordinal variable reflects a continuous latent variable that measures global health and then estimate this latent variable. The first method assuming that self-assessed health follows a lognormal distribution (Wagstaff & van Doorslaer, 1994), the second one using an ordered Probit model and several different dimensions of health to estimate a "health capital" (Cutler & Richardson, 1997) and the last method introducing the use of a health distribution (van Doorslaer & Jones, 2003). In the following subsections, we describe these methods and in the last subsection we discuss the features of our indicator as compared to these three methods.

2.5.1 Getting continuity from an "arbitrary" distribution

When there are no other information on the actual distribution of health, a health measure can be generated by imposing a functional form for its distribution, which relies on empirical observations of the distribution. Wagstaff and van Doorslaer (1994) propose to assume that the observed health distribution over a self-assessed health composed of A categories is generated by a latent unobservable and continuous variable with a standard normal density function. In the course of their analysis, the choice of an inverse lognormal distribution is preferred as regard to the skewed distribution of most of health indicators. Typically persons suffering from serious ill-health are in minority and a large proportion of any general sample population report good health[23]. Indeed, health distributions are strongly concentrated among good health statuses whereas they are spread among lower health statuses, which are more graded (cf. figure 2.6).

Economists often model the distribution of income or wealth using a lognormal distribution (Cowell, 2000). The lognormality has some convenient properties, such as its simple relationship to the normal distribution, the preservation under loglinear transformations as well as the advantage of allowing for skewness. This last point is particularly important for the underlying distribution of health.

The cardinalisation process considers the frequency of each category and calculates thresholds by fitting quantiles from the ordinal categorical variable, notably the cumulated frequencies of categories of self-assessed health, with those of the inverse lognormal distribution. Category scores are obtained as the expected values within each of the intervals defined by the cut points.

[23]The choice of an inverse or a standard lognormal distribution is explained by the skewness of the distribution. If this skewness is observed on the right (respectively left) then an inverse (a standard) lognormal would be preferred.

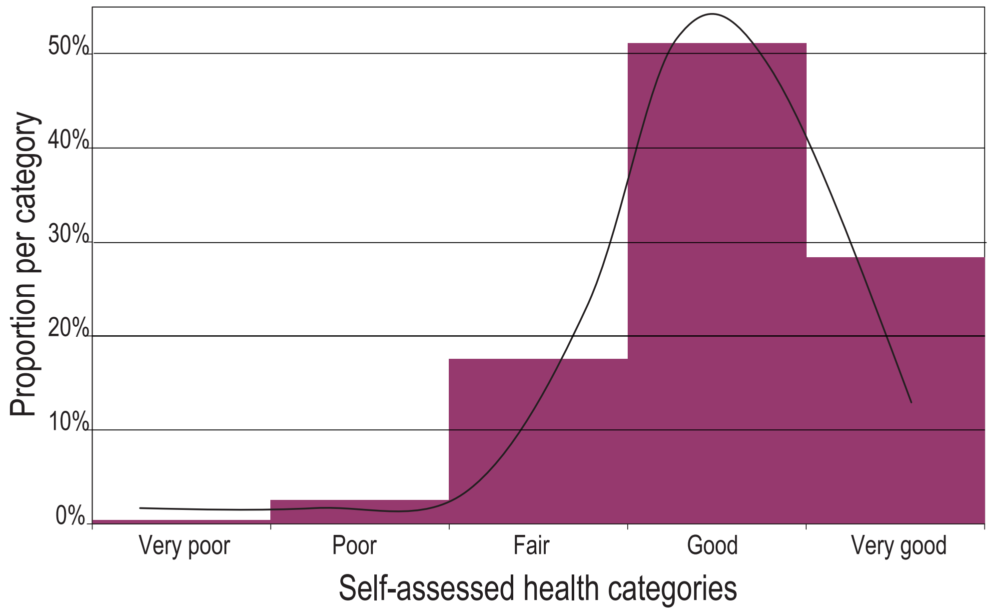

Figure 2.6: The inverse lognormal distribution underlying self-assessed responses

If an individual reports a health status a, his continuous health status is defined by the theoretical average value of the latent health variable between the thresholds c_a and c_{a+1}.

Gerdtham *et al.* (1999) validate this approach. They compare the direct assessment of health status using either the rating-scale method[24] or the time-trade-off[25] method. The main advantage of the Wagstaff and van Doorslaer's approach by comparison to the time-trade-off or to the rating scale, is that categorical information on health status is available in most of the population surveys because this indicator is much easier to collect. However, even if the latent health variable is assumed to be continuous, it is still inherently categorical and therefore it could not be used as a continuous variable in an ordinary least square regression. Its use would produce non normal and heteroscedastic residuals leading to inefficient estimates of coefficients and biased estimates of their standard error. Moreover, intra-categorical differences are not considered. The time-trade-off and the rating scale directly yield a continuous health measure whereas the third method requires an assumption of the shape. This assumption relies rather on arbitrary than obvious feature of the distribution. In particular, it assumes the same distribution of health, whatever the population considered, which may lead to biased estimates of concentration index.

[24]The rating-scale method uses a visual-analogic scale from 0 to 100 with labeled anchors from "death" to "full health".

[25]Individuals are asked to evaluate on a scale of 20, the number of years in full health that they think is of equal value to 20 years in their current health status.

As regard to these critics, a cardinalisation of the self-assessed health using health information in order to overcome the arbitrary aspect.

2.5.2 Getting continuity by combining different health dimensions

Cutler and Richardson (1997) discuss a theoretical framework for measuring health capital of the population. They aim to estimate quality-adjusted life years (QALY), that are weights reflecting the quality of life that somebody attaches to each of his remaining years of life taking into consideration his health conditions during these years. An individual's quality of life is scaled on a 0 to 1 basis, where 0 is equivalent to death and 1 is equivalent to perfect health. Cutler and Richardson (1997) advocate a health measure, which relies not only on a physical measure of morbidity but which accounts also for mental and physical functioning as well as risk factors. Therefore, they choose to estimate QALY by weighting the fact of living with major chronic diseases and functional impairments. This means that suffering from a disease attributes to the individual a quality of life comprised between 0 and 1 (both excluded). Considering the possibility of using time-trade-off methods for the assessment of QALY weights, they reject this approach and argue that "there is no consensus in the literature about the disutility associated with various conditions or the change in these disutilities over time". However, they include a discount rate $\frac{1}{(1+r)^k}$ to take into account individual preference for present.

In this context, using the American National Health Interview Survey, each functional limitation is weighted measuring the extent to which a disease influences self-assessed health. Their method is to assume that people have a latent measure of health, related to their diseases, demographic characteristics and to estimate such a model using an ordered Probit model. The estimated coefficient of the diseases vector is used as a measure of health. The ordered Probit model allows to estimate all the cut-off points of the self-assessed health categories. As a QALY is scaled on $[0; 1]$, the estimated coefficient (usually range from -∞ to ∞) has to be normalised. It is therefore divided by the differences between the estimated coefficients of the highest and the lowest categories of self-assessed health. The estimated coefficient of the diseases vector is interpreted as a reduction in quality of life associated with each chronic condition.

A peculiar aspect is that the QALY loss to a chronic disease is not conditioned by other variables, such as income and standard of living. Indeed, the estimated coefficient of a particular chronic condition informs how that condition changes along the scale of self-assessed health, holding constant demographic characteristics and other reported health conditions. However, a chronic disease has a different impact on an unskilled worker than on a manager, and these aspects are not considered. Indeed, a good utility function must take into account individual preferences in a given context of perfect information, as it is in Grossman (1972) as well as in a given context of uncertainty.

Nevertheless, the validity of this method has not been shown (van Doorslaer & Jones, 2003). Moreover, there is a misspecification of the quality of life; when an individual rates his health as very poor, QALY equals 0, which implies "death" according to preliminary hypotheses whereas the individual is not obviously dead. This construction could lead to give individuals predicted values of health status lower than 0 or greater than 1. Van Doorslaer and Jones (2003) highlight this limitation and offer to overcome it with two alternatives. Firstly, they propose to rescale to the $[0;1]$ interval, using the largest and the lowest prediction. Secondly, under the assumption that a continuous health distribution is available for the sample considered, the range of average values of this distribution for age groups could be used as an explained variable. The minimum and maximum predictions from this new model would then define the observable range of the distribution conditional on the set of regressors.

As regard to critics formulated against these two first methods to describe the latent health variable, a third method proposes to consider a health distribution, in some cases external within an interval regression.

2.5.3 Getting continuity using external information

The third solution relies on the creation of scattering within categories of self-assessed health by considering a health distribution. An appropriate econometric procedure to do so has been proposed by Stewart (1983). It uses a likelihood function for the application at hand. The likelihood function is a modification of that used in the estimation of the standard ordered probit model and replaces the unknown threshold values by the set of known thresholds that delineate the intervals. The responses on the dependent variable are grouped. In the literature this type of model is referred to as a grouped dependent variable model or as interval regression model. As self-assessed health is an ordinal variable in nature but interval coded, this interval nature is exploited within an interval regression model.

In order to understand how the model is implemented, responses of self-assessed health are coded 1, 2,..., 5 to capture the five distinct health status categories. We shall denote y_i the observe self-assessed health and y_i^* an underlying variable that captures the health status of the i^{th} individual. This can be expressed as a linear function of a vector of explanatory variables X_i using the following relationship:

This model can formally be written as

$$y_i^* = X_i\beta + u_i \tag{2.12}$$

It is assumed that y_i^* is related to the observable ordinal variable y_i as follows:

$$
\begin{aligned}
y_i = 1 \text{ if } \ c_0 < y_i^* \leq c_1 \\
y_i = a \text{ if } \ c_{a-1} < y_i^* \leq c_a \text{where } a = 2, \ldots, 4 \\
y_i = 5 \text{ if } \ c_4 < y_i^* \leq +\infty
\end{aligned}
\quad (2.13)
$$

(2.14)

where the c_a for a=1,..., 5 denote the interval boundaries. The exact knowledge of the thresholds allows the likelihood function to be specified in a fairly straightforward manner. The variable y_i^* is best interpreted not as a latent measure but one with a quantitative interpretation. The interval regression thus provides a good alternative to ordered Probit model when the limits of the intervals of the parameter of interest are known. Interval regression has been specifically recommended as an appropriate method for analysing results from contingent valuation studies (Donaldson *et al.*, 1998). It has also been successfully applied by van Doorslaer and Jones (2003) on Canadian data, using a health distribution derived from the Canadian National Population Health Survey (CNPHS), namely Health Utility Index (HUI), to rescale the Canadian self-assessed health available in the same survey[26]. The cumulative distribution function of HUI is used as the benchmark, from which the thresholds defining HUI intervals of each self-assessed health level are derived. In concrete terms, the q^{th} quantile of the distribution of HUI corresponds to the q^{th} quantile of the self-assessed health, which is analogous to the previous inverse lognormal rescaling. In a first step, the cumulative frequency of observations for each category is computed. The second step is then to find the quantiles of the cumulative density function of HUI. Each interval is thus limited by a couple $[c_{a-1}; c_a]$, from which an interval regression can be conducted.

The interval regression thus measures individual probabilities to self-assess a health status between $[c_{a-1}; c_a]$ dependent on a vector of demographic and socioeconomic characteristics. It provides efficient estimated parameters, an identifiable variance of the error term and a definition of the scale of the latent health variable. The values of indicator can be interpreted in terms of health utility because they are obtained by rescaling the latent variable with the distribution of HUI, which is a utility-based measure obtained by a Von Newman-Morgenstern procedure.

This method relies on having a dataset that includes both self-assessed health and a cardinal index of health[27]: in their case the Canadian National Population Health Survey

[26]The self-assessed health question is "*In general, how would you say your health is?*" and the five response categories are *excellent, very good, good, fair* and *poor*.

[27]Van Doorslaer and Jones (2003) go further in their conclusions and propose to use these HUI predicted thresholds to compute an interval regression on self-assessed health, even if the survey does not contain any generic health distribution. We come back on this assumption in chapter 4 and test its validity in the context of an analysis of inequalities in France.

(NPHS), which includes self-assessed and the McMaster health utility index (HUI). This is used to construct a mapping from HUI to self-assessed health on the assumption that there is a systematic relationship between the two measures of health, such that those at the bottom of the distribution of self assessed health will also be those at the bottom of the distribution of health utility. This method cannot be replicated to the French context as we do not have at our disposal a dataset containing both self-assessed health and the questionnaire of the Health Utility Index. Although there is a French version of the HUI, this health utility index is experimental and has been developed on a specific and restricted sample of about fifty children[28] (Le Galès *et al.*, 1999).

2.5.4 Some elements of discussion

From the three previous methods, two aspects appear essential for an appropriate measurement of health status.

Firstly, it is advisable to reach the continuous aspect by using several health factors. For instance, Cutler and Richardson (1997) include physical morbidity, mental and physical functioning and risk factors. Similarly van Doorslaer and Jones (2003) rely on an index of health utility along with self-assessed health.

Secondly, it is important to consider the strong links between health and individual characteristics as it is done in ordered Probit as well as in the interval regression, which includes various individual characteristics.

Our construction encompasses these two elements.

Nevertheless, it is noteworthy that the main difference between our procedure and the constructions proposed in Wagstaff and van Doorslaer (1994) or van Doorslaer and Jones (2003) is the initial element of the measurement of health. These methods rely firstly on self-assessed health whereas our initial element is the reported diseases count that we assume more objective as we correct it using a severity index. Then, these methods use a distribution (arbitrary or representing health) assumed more objective to correct the subjective health whereas we rely on self-assessed health to weight the number of diseases. In simple terms, we could say that these methods generate a subjective indicator of health corrected with objective health information. On the contrary, we generate an objective indicator of health corrected with subjective health information. As a result, we all propose a mixed indicator of health but with different initial assumptions.

By comparison to the measure of health proposed by Cutler and Richardson (1997), our indicator is more informative than an indicator that would be based on the occurrence of the disease, because it takes into consideration the fact that some diseases affect the

[28]The French Health Utilities Index has been tested on a particular population. The self-reported questionnaire has been adapted and validated in a population of children with cancer, a group of 42 children already included in a multi-center database designed by the Group on Brain Tumours in Children of the French Society for Paediatric Oncology.

length of life as well as its quality. Moreover, we can underline that our indicator could also easily involve a parameter of preference for present or preference for certainty as proposed in Cutler and Richardson (1997).

2.6 Conclusion

In view of the multidimensional nature of health status and the need to take into account reporting biases, we consider the construction of a health status variable encompassing the three main dimensions of health described by Blaxter (1990), namely medical, functional and subjective, while offering a cardinal health indicator. Firstly, the medical and functional dimensions are translated into the number of diseases and their respective severity level medically evaluated. Secondly, the subjective dimension is approached by self-assessed health level. Despite the fact that diseases are self-declared data and so, can suffer from individual reporting bias, this health information seems to be less biased than self-assessed health because of the use of diseases' severity level. These severity levels allow checking for coherency between severity and number of diseases.

Our model uses both an ordered Probit and new explanatory variables. As a result, the measurement of health we propose is cardinal as it initially relies on a cardinal numeral determinant: the individual number of diseases.

This method gives a simple way to construct a continuous indicator with variables classically collected in health surveys. Moreover, this method could be replicated on previous versions of the survey and it would enable us to study changes over time. It could also be applied with minor adaptations to other surveys as the severity index that we propose is related to the International Classification of Diseases ICD-10. This aggregation and bias correction method could also easily be used with other sociodemographic, health and health related behaviour variables. The main strength of this method is to use retrospective information from health surveys.

Another important result of our study is the significance of the cluster effect due to unobserved heterogeneity among households. It means that important common unobserved factors among households affect either the general health status or the scale itself. We have chosen to use a random effect model to correct this bias. In the process, we found evidence of instability in the value of the coefficients and their standard errors, which reduces their significance. Although the use of a random effect model rather than a fixed effect model is debatable, it is important to stress that if we do not take into account this household effect, it may generate biases, reduce the accuracy of estimates and make coefficients less comparable among populations because of shrinkage. As this household effect is significant for French data, it might also be observed in other countries. However, as far as we know, no studies have considered this household effect in health reports so far.

As to shifting thresholds, their introduction does not substantially modify the values of coefficients associated to the degree of severity, except for the highest one. A model with varying thresholds is more informative; however, in our study such a model does not involve a significant improvement of the estimation and is also costly in terms of time calculation. The model with fixed thresholds is preferred because our main purpose is to use the estimated parameters as weights of the number of diseases to construct a health indicator.

Another benefit of this indicator is that it allows health status comparisons between different populations and distributions analyses. It offers news prospects of analyses such as inequality analysis, which is broached in chapter 4. Furthermore, this indicator could be used as an explanatory variable in other analyses, such as inequalities in health care consumption to define individual need for care. It would permit avoiding collinearity between several explanatory health variables and produces a more parsimonious model.

Part II

Measuring inequalities in health

Chapter 3

Measures of health inequality

3.1 Introduction

Since the pioneering works of Kolm (1969), Atkinson (1970) and Sen (1973), inequality measurement is often based on explicit social welfare function[1]. However inequality measurement relies on the distribution of other dimensions rather than the distribution of welfare itself, and we assume that the latter is derived from the former. What other dimensions are to be taken into account is open to debate. Income is the most widespread dimension but some other factors such as health, handicap, age, or family size (Foster & Sen, 1997; Sen, 1973) are also recommended. A comprehensive literature has dealt with the analysis of income inequality. Indeed, the literature is better documented on the analysis of inequality in income than in any other variable (Trannoy, 1999). According to Deaton (2001), a good way to answer the question of the measurement of inequality in health would be to start with inequality in income and to ask whether the theoretical and the measurement structures of inequality in income could be transferred to inequality in health. The literature on the measurement of income inequality firstly considers a single-dimensioned case, and the Lorenz dominance criterion has been provided in this context (Atkinson, 1970). Nevertheless, as Trannoy (2006) underlines, a consensus has been

> "emerging among many scientists, particularly development economists, about the multidimensional aspect of individual well-being which cannot be reduced to a unique monetary unit".

Furthermore, Foster and Sen (1997) underline that an exclusive concentration on inequalities in income distribution cannot be adequate for an understanding of economic inequality.

[1]Inequality judgements are often, but not always, based on a welfarist concept. Sen (1980) argues that for many purposed, the appropriate context in which to judge inequality is neither that of utilities (as claimed by welfarists) nor that of primary goods (as claimed by Rawls (1982)) but depends on the goal. If the goal is to concentrate on the individual's opportunity to pursue his objectives, then account would be taken not only on primary goods the person holds, but also on the relevant personal characteristics that govern the conversion of primary goods into the person's ability to achieve his goals. For instance, the approach of equality of opportunity which is used in chapter 5, does not require welfare-based measures.

Income is only one factor among many other factors that influence the real opportunities people enjoy. For this reason, the measurement of inequality in income has been extended to a multi-attribute context; taking into account that individual well-being is based on several attributes such as education and health. Therefore, the main concept of univariate analysis and related criteria have been generalised to a multi-variate inequality.

This chapter precedes the empirical analysis of this dissertation. It gives an overview of methods used for measuring inequality in the literature and sets criterion and indices which will be used. When measuring inequality, we aim to achieve two goals. The first goal is to judge whether a particular distribution is more equal than another, and so we concentrate our discussion on methods allowing a ranking of distributions. The second goal is to quantify the difference in inequality between distributions and inequality indices that are used for it.

As a result, this chapter is divided into four sections. The first section answers the question of ranking alternative distributions and summarises the main results on this aspect in the literature of inequality measurement. In particular, we first consider health as a "lonely" attribute of individual welfare. We then add income to the individual welfare and recall principles of the measurement of inequality within a multidimensional perspective. The second and the third sections cover the empirical aspects of the measurement of inequality and they provide a panel of measurement tools based on ranking distributions. The second section proposes a unidimensional context for the measurement of inequality in health. It describes inequality indices. We distinguish these indices by the type of data of health status they require. The third section concerns the approach of income-related inequalities in health and describes inequality indices that take into account two dimensions such as health and income. In particular, the concentration index is comprehensively discussed. In the fourth section, we outline our conclusions.

3.2 Measurement of inequality in health: orderings and rankings

The first way to demonstrate objectively the existence of inequalities in health is to obtain a ranking of distributions in order to compare them. This section draws up a framework for the analysis of inequality. Firstly, we assume that a social welfare function is composed of a single attribute. We then move toward a bivariate approach of inequalities where the social welfare function is assumed to depend on two attributes. In this context, alternative cases on the role played by dimensions are considered. Firstly, the two attributed are considered symmetrical. Secondly, a property of transferability links the two attributes so that one attribute can be used to compensate for a deficiency in the other one.

3.2.1 Stochastic dominance: first and second order

We rely on Sen (1973), who comprehensively discusses the measurement of inequality in a univariate context, and summarise the usual criteria of dominance. In a unidimensional context, we assume individual utility to be defined on a lonely attribute, which, for convenience, we shall refer to as health[2]. In general, health can be either dichotomous, such as suffering from a disease or not, or ordinal, such as the self-assessed health in five categories, or even cardinal, such as quality adjusted life measure (QALY). We shall assume a population composed of n individuals i such that $i = 1, \dots, n$.

Assumption 1: Health is a qualitative variable

In a first step, we assume that health is a qualitative variable composed of k alternative health statuses, i.e k is the number of categories of the qualitative variable such as $x = \{x_1, x_2, \dots, x_k\}$. These k health statuses are ordered[3] according to $x_1 \leq x_2 \leq \dots x_{n-1} \leq x_k$, where the most favourable health status is x_k. Comparisons between distributions can be done using social welfare functions. We assume a social welfare function, which is a function of individual utilities. In order to formulate any ranking of distributions, we have to make some assumptions about the form of these individual utility functions $U(x)$. In a first step, we restrict ourselves to the class of functions $U(x)$ which are increasing. We consider on which conditions we can rank two health distributions within this restricted frame. From this ordinal health status indicator, a stochastic dominance criterion at first order can be proposed.

> Definition 1: *Stochastic Dominance at First Order*
> *Given any two health distributions x and x', with respective cumulative distribution functions $F_x(x)$ and $F_{x'}(x)$, we say that x dominates at first order x', written $x \geq_{SD_1} x'$, if and only if $F_x(x_j) \leq F_{x'}(x_j)$, for any health status $x_j = \{x_1, x_2, \dots, x_k\}$*

Stochastic dominance at first order displays a weak condition for welfare ranking. Indeed, it means that health is better in distribution x than distribution x' for each category of health status: the share of the population in the worst category of health is lower (or no higher) for x than x' as well as the share of the population in the lowest two categories, the lowest three categories, and so on. Stochastic dominance at first order means that whatever the health status considered x_j, the probability to get a health level equal or

[2] The conventional approach is to assume that the first attribute is income.

[3] We assume an ordering of the alternatives that involves a ranking with two properties: completeness and transitivity. The property of completeness requires that for any pair of alternative health statuses x_{j_1} and x_{j_2}, either $x_{j_1} \Re x_{j_2}$ holds or $x_{j_1} \Re x_{j_2}$, or both. If we assume that the ranking relation $\Re$ means "at least as good as". In this context, if $x_{j_1} \Re x_{j_2}$ holds and not $x_{j_2} \Re x_{j_1}$ then x_{j_1} is strictly better than x_{j_2} and conversely. If both $x_{j_1} \Re x_{j_2}$ and $x_{j_2} \Re x_{j_1}$ hold then x_{j_1} and x_{j_2} are indifferent. The property of transitivity relies on three alternative statuses x_{j_1}, x_{j_2} and x_{j_3} and implies that if $x_{j_1} \Re x_{j_2}$ and $x_{j_2} \Re x_{j_3}$ both hold then so does $x_{j_1} \Re x_{j_3}$.

higher than x_j is always lower with distribution x' than with the distribution x. Therefore, any rational agent prefers x to x' and stochastic dominance at first order of distribution x over x' implies that any social welfare function which is increasing in health will record higher levels of welfare in x than in x'.

Stochastic dominance at first order is not sensitive to inequality. A key question in the measurement of inequality is how to interpret the distributional information imparted by such qualitative data. In order to use standard measurement methods from income inequality analysis, various techniques could be proposed to use qualitative data. A first standard technique is to assign a numerical value to each category of health. For instance, if we consider self-assessed health in five categories[4], it consists in assigning the values 1 to 5 to the respective categories very poor to very good. Alternatively, another technique is to impose a specific scale allowing differences between categories. A third technique relies on the measuring inequality using mean as traditional measures of inequality are all mean-based. In this context, inequality is measured as a deviation from the mean. These techniques are not fully satisfactory but still most of the published empirical analyses use Gini-based measures.

Allison and Foster (2004) carry out extensive review of these methods. They particularly underline the non-robustness of the mean both as an indicator of societal health and as a reference point. As regard to these limits, they propose an alternative reference point, based on the median. They argue that there is a natural centre at the 50^{th} percentile in a population health status variable. The median is always located at the position where half the population has a health status below (or equal to) it and half above it. Unlike the mean, the relative position of the median does not change as the scale changes. This median-based approach views inequality as a spread away from the median as defined in the following definition.

> Definition 2: *Spread Away from the Median Ranking*
> *Given any two health distributions x and x', with respective cumulative distribution functions $F(x)$ and $F'(x)$ and median state $m(x)$ and $m'(x)$ distribution x health has greater spread than distribution x', written xSx', if x and x' have the same median category $m(x) = m'(x) = m$;*
> *for all $k < m$, $F_k(x) \geq F'_k(x)$*
> *for all $k \geq m$, $F_k(x) \leq F'_k(x)$*

In other words, distribution x' dominates at first order distribution x below the median while distribution x dominates at first order distribution x' for the median category and above. Therefore the ranking relationship S is a partial ordering of distributions, a $S-$curve can be represented from the partial ordering S (cf. Graph 3.1). The construction of the $S-$curve is based on a traditional representation of the cumulative distribution

[4]We refer to the five typical categories for self-assessed health variables, namely "very poor, poor, fair, good and very good".

function (CDF) of a qualitative variable, from which the portion of the CDF situated on the right of the median is flipped over to the left. The resulting curve is then rotated 90^o such that the base of the $S-$curve represents the range of the population having the median level of health. The height of the $S-$curve is thus the number of health categories on the right and on the left of the median.

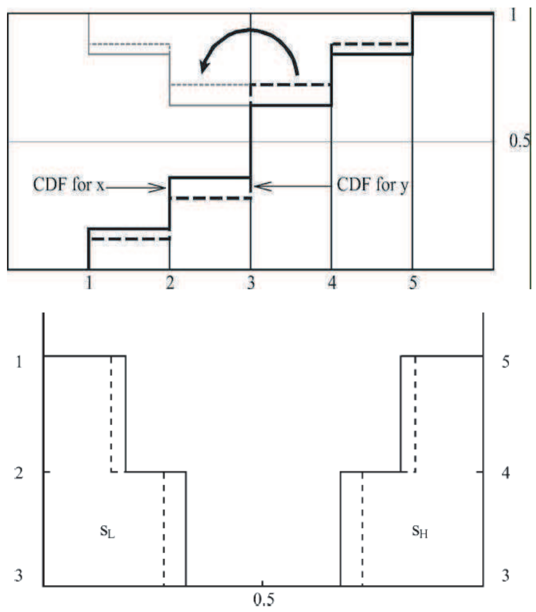

Figure 3.1: The S-curve (*source: Allison & Foster (2004)*)

A higher $S-$curve indicates greater inequality according to the ranking S. Conversely, if the $S-$curves cross, then the two distributions cannot be compared using S. Allison and Foster (2004) thus introduce the following theorem.

> Theorem 1: *Ranking according to the $S-$curve*
> *Given any two health distributions x and x', having the same median category*
> *xSx' if and only if the $S-$curve of x is no lower than the $S-$curve of x'.*

Associated with the graphical representation of the $S-$curve are several basic measures of spread which are helpful in signalling and explaining changes in inequality in health across states. We shall define $s(x, c)$ as being twice the area below the $S-$curve of distribution x, with c the scale of the qualitative variable of health. We denote $s_L(x, c)$ as twice the area to the left of 0.50, the median and $s_H(x, c)$ as twice the area to the right. Consequently, $s_L(x, c)$ is the average number of categories that the lower half of the distribution falls below the median, while $s_H(x, c)$ is the average number of categories that the upper half is above the median. Allison and Foster (2004) call these lower spread and upper spread respectively and define the following equality.

Definition 3: *Lower spread and upper spread*
Given $s_L(x,c)$ *and* $s_H(x,c)$ *the average number of categories that the lower (resp. the upper) half of the distribution falls below (resp. above) the median, the spread index* $s(x,c)$ *can be written*
$s(x,c) = s_L(x,c) + s_H(x,c)$

The partial ordering S is equivalent to an unambiguous ranking by the spread index $s(x,c)$ across all scales. Moreover, as the spread index $s(x,c)$ aggregates the partial indices $s_L(x,c)$ and $s_H(x,c)$, these two indices provide a greater comparison between two distributions. They permit disentangling changes below the median from those above the median. Theorem 5 extends this property.

Theorem 2:
Given any two health distributions x *and* x'*, having the same median category* xSx' *if and only if* $s_L(x,c) \geq s_L(x',c)$ *and* $s_H(x,c) \geq s_H(x',c)$, $\forall c$.

This theorem is empirically useful as it relies on the computation of s_L and s_H. If these two partial spread indices order distributions in the same direction, then a ranking by S is possible. If they disagree, then a ranking by S is impossible. Nevertheless, this disagreement is indicative of stochastic dominance at first order[5]. The main advantage of this methodology is to permit measuring inequality independently of a scale. Moreover this measurement of inequality does not require any transformation of the ordinal variable such as cardinalisation. The nature of the self-assessed health variable remains unchanged and there is no arbitrariness involved in the construction of a usable health variable. The price of these advantages is to propose an analysis framework restricted to situations where medians of distributions coincide. Furthermore, usual criticism concerning the individual's reporting heterogeneity in self-reported variables (we refer to chapter 1).

Apouey (2007) provides an interesting empirical application of this method using self-assessed health. She stresses the common ground between the axiomatic foundation of dispersion measures and the median-based polarisation measures of income, which are an alternative to mean-based inequality measures. Her analysis compares polarisation indices for self-assessed health on the basis of cumulative distribution functions and Gini indices. The alternative measures of inequality are shown to address separate questions, and so are complementary. To our knowledge, there are few empirical papers (Apouey, 2007; Abul Naga & Yalcin, 2007) which follow the method proposed by Allison and Foster (2004) whereas many researchers look into the cardinalisation or the reliability of self-reported health data in inequality in health analyses (see most of the literature references in chapter 2).

[5] The corresponding theorem is Theorem 3:
Given any two health distributions x *and* x'*, having the same median category* $x \geq_{SD_1} x'$ *if and only if* $s_L(x,c) \leq s_L(x',c)$*, and* $s_H(x,c) \geq s_H(x',c)$, $\forall c$.

Allison and Foster (2004) present a median based rating robust to different cardinal scales used for health status (Zheng, 2006). Their median-based partial ordering is introduced in the context of first order dominance, which yields unambiguous orderings of aggregate or mean health. If the stochastic dominance criterion at first order does not permit an unambiguous ranking of distributions then we use the stochastic dominance criterion at second order. Mathematically, stochastic dominance at first order implies stochastic dominance at second order. We shall now assume that a health indicator suitable for the construction of the distributional analysis index is available. This assumption is tenable as health can be either established as cardinal with specific survey questionnaires such as the Health Utility Index or transformed into a continuous and cardinal health indicator as illustrated in chapter 2.

Assumption 2: Health is a cardinal variable

We assume a set of health statuses represented by real numbers in a population composed of n individuals. We have to make some assumptions about the form of the utility function $U(x)$, which depends on individual health only. We restrict ourselves to the class of functions $U(x)$ which are increasing and concave. These assumptions are reasonable because a supplement unit of health increases individual utility at a decreasing rate.

Although stochastic dominance at second order relies on stronger assumptions than stochastic dominance at first order, this second level may be required for two reasons (Cowell, 2000). In practical applications, it is very often the case that neither distribution dominates another one at first order. Secondly, stochastic dominance at first order does not employ all the standard principles of social welfare analysis: above all, it does not incorporate the transfers principle.

This transfers principle is important in most of the inequality literature. It expresses that inequality is certain to be diminished by a series of transfers such that all transfers from a richer individual to a poorer individual still leave the former individual richer than or just as rich as the latter individual (Pigou, 1920; Dalton, 1920). These transfers are said to be progressive. The plausibility of the transfers principle for health has been discussed by Bleichrodt and van Doorslaer (2006) and we will discuss this point in the end of this section. At present, we assume that the health transfers principle holds if a transfer of health from someone who is in better health to someone who is in worse health does not lead to a reduction in social welfare, provided that the transfer does not change the health ranking of individuals. Therefore, we write the Pigou-Dalton health transfers principle as follows.

Definition 4: *Pigou-Dalton transfers principle:*
Given any two health distributions x and x', we say that x dominates in the Pigou-Dalton sense x', written $x \geq_{PD} x'$, if x can be obtained from x' by a finite number of progressive transfers.

We now introduce the stochastic dominance criterion at second-order, which relies on comparisons of integrals of the cumulative distribution functions (CDF) of two health distributions[6].

Definition 5: *Stochastic Dominance at Second Order*
Given any two health distributions, distribution x dominates at second order distribution x', written $x \geq_{SD_2} x'$, if and only if
$\int_0^k x(t)dt \leq \int_0^k x'(t)dt$, $\forall k = 1, ..., n$.

If the integral of the CDF of distribution x lies nowhere above and somewhere below that of distribution x' then distribution x dominates distribution x' at second order. Therefore, welfare function will thus record higher levels of welfare in x than in x'.

The stochastic dominance at second order is equivalent to Lorenz dominance. The Lorenz dominance allows us to make a partial ranking of distributions without knowledge of the precise form of the social welfare function (Atkinson, 1970). The Lorenz curve gives an interpretation of inequality using health share without considering differences in mean health. We draw the Lorenz curves corresponding to a health distribution x in $[0, 1]^2$, the unit square.

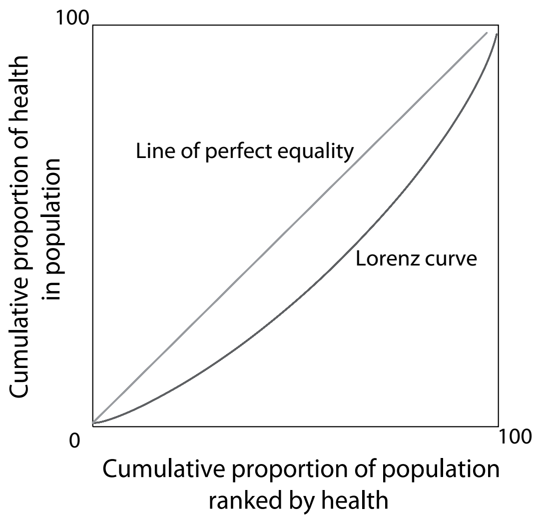

Figure 3.2: The Lorenz curve applied to health

[6]In order to improve inequality measurement, we have imposed the availability of a cardinal variable prior to the definition of stochastic dominance at second order.

The principle of the Lorenz curve is to observe the deviation of each health level[7] from the heath share that corresponds to perfect equality, namely the diagonal of that square. As a result, the farther the curve falls below the diagonal, the higher inequality is. When the Lorenz curve coincides with the diagonal, then there is no inequality. The figure 3.2 gives a graphical representation of a Lorenz curve.

Inequality indices have been proposed from this graphical aspect of the Lorenz curve, namely the Gini index that we will present later. The partial ranking of distributions according to Lorenz relies on the comparison of Lorenz curves and is defined as follows.

> Definition 6: *Lorenz Dominance*
> *Given any two ordered health distributions x and x', distribution x Lorenz dominates distribution x', written $x \geq_{\mathrm{L}} x'$, if the Lorenz curve of x lies everywhere above that of x'.*

We then say that there is less inequality in distribution x than in distribution x'. When two Lorenz curves overlap, the partial ranking can no longer be applied.

From a welfare point of view, an important result for the measurement of inequality is the Hardy-Littlewood-Polya theorem (Hardy *et al.*, 1952) which brings all the previous conditions together in a unifying manner.

> Theorem 4: *Hardy-Littlewood-Polya (1952)*
> *Given any two health distributions x and x' of same mean, such that $x_1 \leq x_2 \leq ... \leq x_n$ and $x'_1 \leq x'_2 \leq ... \leq x'_n$, the three following conditions are equivalent:*
> *(i) distribution x dominates at second order x'*
> *(ii) distribution x dominates x' in the Pigou-Dalton sense*
> *(iii) distribution x Lorenz dominates x'*

The interest of this theorem[8] comes from the fact that it establishes the equivalence not only between Pigou-Dalton transfers and Lorenz dominance but also between these notions and welfare dominance (Trannoy, 2006). It has been called the fundamental theorem of inequality measurement(Ok & Kranich, 1998).

Up to now, we have described the measurement of inequality within an analysis framework where we consider individual welfare only based on health. Nevertheless, the unidimensional approach presents some drawbacks.

The first drawback concerns the unidimensional aspect itself. Indeed, the definition of individual welfare on a lonely attribute is restrictive. Many economists support a definition of welfare, which relies on several individual determinants, such as income, health, education, etc. There is a need to supplement the initial attribute, namely income by

[7]Property of cardinality that we have previously assumed is particularly important to draw the Lorenz curve.

[8]For extensive discussion and alternative proofs of this theorem, we refer to Rothschild and Stiglitz (1973) and Sen (1973).

other attributes of well-being, such as health and education. Sen (1973; 1982) also suggests attributes other than welfare such as the various things a person may value doing or being, that he calls the space of "functionings".

The second drawback concerns the specific unidimensional approach of inequalities in health for which the acceptability of the Pigou-Dalton transfers principle for health has to be discussed. In this specific context, the principle means that transferring health from someone with higher health to someone with lower health does not lead to a reduction in social welfare provided the transfer does not change the health ranking of individuals. Nevertheless, it is not always desirable to transfer health from a healthier individual to a less healthy individual, especially when the healthier person is poor and the less healthy individual is rich[9]. Bleichrodt and van Doorslaer (2006) discuss the plausibility of this principle. They argue that it is more acceptable if the correlation between health and income is particularly strong. They therefore recommend to study multivariate concept of inequality in order to care both about the distribution of health and about the distribution of other attributes. As a consequence, it is of interest to extend the analysis framework to a bi-dimensional context. We consider income as a second attribute defining individual well-being.

3.2.2 Multidimensional welfare analysis: symmetrical attributes

The literature on multidimensional analysis extends various criteria developed in the unidimensional framework. Nevertheless, conversely to unidimensional context, evidence on equivalence between these criteria has not been given and may not exist. Therefore, the derivation of social dominance conditions for multidimensional welfare analysis is one of the main challenges of modern welfare analysis.

Following the previous section, we assume that when the inequality of health is represented by a distribution, this distribution describes not only health but also differences in health status due to a number of other relevant attributes, such as age, gender, income and various socioeconomic conditions. For simplicity's sake, we limit the multidimensional context as a bidimensional context and consider the bivariate distribution of a vector $x = (x_1, x_2)$. The situation is clearly visualised if x_1 represents income and x_2 is health.

Income as well as health can be either qualitative, such as quintiles or cardinal, such as an equivalent household income. A bivariate allocation of health and income can thus take three forms :

(i) The two attributes are ordinal variables

[9]Similarly, a socioeconomic version of the principle of health transfers (Bleichrodt & van Doorslaer, 2006) which consists in transferring health from someone who is socially richer to someone who is socially poorer, provided that the transfer does not change the ranking of the individuals in terms of socioeconomic status, is also contestable. Indeed, it does not seem desirable to transfer health from a person socially advantaged to a person socially disadvantaged when the former is in poor health.

(ii) The two attributes are cardinal

(iii) One attribute is ordinal and the other one is cardinal. This last case will be considered later in the context of multidimensional welfare analysis concerning asymmetrical attributes.

We shall assume a population composed of n individuals i such that $i = 1, \ldots, n$. We assume that the vector x is defined in $[0, a_1] \times [0, a_2] = A_1 \times A_2$ where a_1 and a_2 are in $\mathbb{R}^+$. We denote the corresponding joint cumulative distribution function by $F(x_1, x_2)$ and the cumulative distribution function of x_1 and x_2 by respectively $F_1(x_1)$ and $F_2(x_2)$. We assume two bivariate distributions x and x', respectively represented by two cumulative distribution functions $F(x_1, x_2)$ and $F'(x_1, x_2)$.

In this context where attributes are considered as symmetrical, the results of Atkinson and Bourguignon (1982) are important to consider. Their aim is to to seek unanimity among social welfare functions over the ranking of allocations. The comparison of two bivariate distributions is based on the difference in expected utility. The results on stochastic dominance for the multivariate case and *a fortiori* for the bivariate case concentrate on stochastic dominance at first-degree.

Assumption 1: The two attributes are ordinal variables

We shall consider specific hypotheses on the individual utility functions.

Following Atkinson and Bourguignon (1982), we define the following classes of individual social welfare functions:

1. $\mathcal{U}^-$ represents individual utility continuous and increasing in each attributes for which cross-derivative is negative:
   Class $\mathcal{U}^- : U_1, U_2 \geq 0; U_{12} \leq 0, \forall x_1, x_2$

2. $\mathcal{U}^+$ represents individual utility continuous and increasing in each attributes for which cross-derivative is positive:
   Class $\mathcal{U}^+ : U_1, U_2 \geq 0; U_{12} \geq 0, \forall x_1, x_2$

Multivariate dominance criteria at first and at second order rely on these large classes of social welfare functions, which are defined by the signs of their derivatives. We firstly present multivariate stochastic dominance criteria at first order (Atkinson & Bourguignon, 1982), their definition relies on the joint cumulative distribution function, $F(x_1, x_2)$ and the expression

$$K(x_1, x_2) = -[F(x_1, x_2) - F_1(x_1) - F_2(x_2)] \tag{3.1}$$

We can then define dominance criteria at first order in the bidimensional approach according to the following theorem[10].

Theorem 5: *Stochastic Dominance at First Order*
Given any two distributions X and X',

1. *For all individual utility functions $U \in \mathcal{U}^-$*
   $X \geq_{SD_1} X'$ *if and only if* $\Delta F(x_1, x_2) \leq 0$, $\forall x_1, x_2$,
2. *For all individual utility functions $U \in \mathcal{U}^+$,*
   $X \geq_{SD_1} X'$ *if and only if* $\Delta K(x_1, x_2) \leq 0$, $\forall x_1, x_2$,
3. *For all individual utility functions $U \in \bigcup(\mathcal{U}^-, \mathcal{U}^+)$.*

If the marginal distributions are identical then conditions 1 and 2 can only be satisfied simultaneously if $\Delta F(x_1, x_2) = 0$ for all x_1, x_2. The stochastic dominance criterion at first order then relies on correlation between the two attributes involved in the utility function. If x_1 and x_2 are independent, then conditions are reduced to those for dominance at margins. In this context, multivariate stochastic dominance at first order is equivalent to univariate stochastic dominance at first order applied to the two marginal distributions. Stochastic dominance at first order requires strong conditions on distributions. When we cannot unambiguously conclude that there is dominance at first order, we extend the analysis to stochastic dominance at second order.

Assumption 2: The two attributes are cardinal variables

The stochastic dominance at second order assumes additional hypotheses on individual utility functions and relies on classes $\mathcal{U}^{--}$ and $\mathcal{U}^{++}$.

1. $\mathcal{U}^{--}$ represents individual utility functions from $\mathcal{U}^-$ and is such that third derivatives are positive and fourth derivative is negative:
   Class $\mathcal{U}^{--}$: Conditions for $\mathcal{U}^-$ and $U_{11}, U_{22} \leq 0; U_{112}, U_{122} \geq 0; U_{1122} \leq 0, \forall x_1, x_2$

2. $\mathcal{U}^{++}$ represents individual utility functions from $\mathcal{U}^-$ and is such that third derivatives are negative and fourth derivative is positive:
   Class $\mathcal{U}^{++}$: Conditions for $\mathcal{U}^+$ and $U_{11}, U_{22} \leq 0; U_{112}, U_{122} \leq 0; U_{1122} \geq 0, \forall x_1, x_2$

[10] Condition 1 is a theorem proposed by Hadar and Russell (1974). They also show that it can be extended to the multidimensional context. Condition 2 is a theorem that comes from Levy and Paroush (1974).

We shall assume the following expressions:

$$H(x_1, x_2) = \int_0^{x_1} \int_0^{x_2} F(s,t)dtds \tag{3.2}$$

$$H_1(x_1) = \int_0^{x_1} F_1(s)ds \tag{3.3}$$

$$H_2(x_2) = \int_0^{x_2} F_2(s)ds \tag{3.4}$$

$$L(x_1, x_2) = \int_0^{x_1} \int_0^{x_2} K(s,t)dtds \tag{3.5}$$

The expressions 3.4 and 3.5 are related to the marginal distributions of each attribute whereas the expressions 3.3 and 3.5 rely on the joint distribution of the two attributes.

From these expressions, Atkinson and Bourguignon (1982) define stochastic dominance at second order in a bidimensional approach and propose the following theorems.

Theorem 6: *Stochastic Dominance at Second Order*
Given any two bivariate distributions X and X'

1. *for all individual utility functions $U \in \mathcal{U}^{--}$,*
   $X \geq_{SD_2} X'$ if and only if
   $\Delta H_1(x_1) \leq 0$, $\Delta H_2(x_2) \leq 0$, $\Delta H(x_1, x_2) \leq 0$, $\forall x_1, x_2$
2. *for all individual utility functions $U \in \mathcal{U}^{++}$,*
   $X \geq_{SD_2} X'$ if and only if
   $\Delta H_1(x_1) \leq 0$, $\Delta H_2(x_2) \leq 0$, $\Delta L(x_1, x_2) \geq 0$, $\forall x_1, x_2$

When attributes are independent, then the criterion is equivalent to a unidimensional framework. Nevertheless, we know that x_1, which stands for health, and x_2, which stands for income, are dependant as shown by many empirical studies (e.g Deaton, 2003; Wagstaff & van Doorslaer, 2000). In this latter context, Atkinson and Bourguignon (1982) emphasise an interpretation of the dominance conditions in terms of incomplete covariance, which can be compared to the interpretation in terms of incomplete means, such as Lorenz curve in the unidimensional case. Concerning means of the two distributions, if they are the same, the distribution with the higher covariance cannot dominate the other distribution and reversely the distribution with the lower covariance cannot dominate either. When means differ, a distribution with higher means can dominate.

Distinctive features of the bivariate allocation of income and health

Atkinson and Bourguignon (1982) provide tools to compare bivariate distributions. An illustration for health and income is the comparison of distributions of income and life expectancy at two different dates. The issue is therefore to show whether the distribution in one year can dominate the distribution in another year for all social welfare functions in one of the classes identified earlier. The comparison thus relies mainly on marginal conditions for each attribute and changes in sign of $\Delta L(x_1, x_2)$ and $\Delta H(x_1, x_2)$, the two attributes assuming to play a symmetrical role.

This approach is essentially an extension of stochastic dominance criteria in a unidimensional framework. Conditions on individual utility functions are therefore a technical extension, whose intuitions are not straightforward. In particular, the meaning of the cross fourth partial derivative is difficult to understand (Le Breton, 1986). Moreover, unlike unidimensional context, Atkinson and Bourguignon's approach does not rely on a transfers principle[11].

Furthermore, the two attributes that are income and health may be either complement or substitutable. In the complementary case, individuals' preferences rely on the correlation between the two attributes. Therefore, health and income are playing a symmetric role, as illustrated in this section. In substitutability context, one of the two attributes can be used to compensate for a deficiency in the other one. In the income-health example, transfers of income to compensate poor health is easily conceivable. We shall now consider an asymmetrical treatment of attributes, which offers relevant perspectives to measurement of inequality and allows extensions of notions from unidimensional dominance.

3.2.3 Multidimensional welfare analysis: asymmetrical attributes

Until now, the two attributes composing individual utility have been considered as playing a symmetrical role. Nevertheless, Bommier and Stecklov (2002) argue that social welfare is inconsistent with the basic notion of a just or equitable distribution of health if the utility function is based on two arguments with symmetric roles. Indeed, whatever restrictions made on the cross derivatives of the utility function, the social welfare approach does not reject an income-based discrimination in access to health: either there is discrimination in favor of the poor, even if health is not related to income or there is discrimination in favor of the rich. As a consequence, it appears to be more appropriate to consider a situation where the two attributes are asymmetrical[12]. Muller and Trannoy (2003) introduce an asymmetrical treatment of attributes, which relies on a compensation principle. Their main argument is to consider that among all the attributes at least one can be used to make direct transfers between individuals. When the two attributes we refer to are income and health, the compensation principle can be considered. Indeed, the sickest are often also the poorest and compensation with income is advised. Their analysis is in the continuation of Atkinson and Bourguignon's results on symmetrical attributes and on the needs approach. We shall consider the results of the need approach

[11] The transfers principle has been introduced by Moyes (1999).

[12] Bommier and Stecklov (2002) do not consider this alternative and concentrate on an alternative approach, that they consider more appropriate for equity principles. They follow Rawls' principles of justice and provide a measure of the distance between the actual and the ideal distribution of health. Their approach is closely related to the concentration index that will be presented in section 3.4.2.

before summarising the asymmetrical treatment of attributes as proposed by Muller and Trannoy (2003).

Assumption 1: One of the attributes is an indicator of needs

The needs approach is a way of extending dominance results to cases where individuals differ in needs as well as in income. Atkinson and Bourguignon (1987) propose a robust approach, which is based on an ordinal classification of all the households into different need groups. If we transfer their approach to a more general context, the first attribute is used to divide the population in homogeneous subgroups and distribution of the second attribute is represented within the groups and the whole society. This approach is appropriate when the first attribute refers to income and the second attribute refers to health. If we refer to the previous subsection, we consider the third possible form of the pair of variables : one of the attribute is ordinal and the other one is cardinal.

Their approach relies on the sequential generalized Lorenz dominance (SGLD) criterion which is based on the generalised Lorenz criterion. The generalised Lorenz curve (GL) is obtained by scaling up the ordinary Lorenz curve by the mean of the distribution and so, plots cumulative shares of the variable of interest scaled by the mean of the distribution against cumulative population. The corresponding ranking has been implemented by seeking a dominance relation between generalized Lorenz curves (Shorrocks, 1983).

> Definition 7: *Generalised Lorenz Dominance*
> *Given any two ordered health distributions x and x', such that $x_1 \leq x_2 \leq ... \leq x_n$ and $x'_1 \leq x'_2 \leq ... \leq x'_n$,*
> *distribution x dominates in the generalised Lorenz sense distribution x', written $x \geq_{\text{GL}} x'$, if*
> $\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} x'_i, \forall k = 1, ..., n.$

As for the sequential generalised Lorenz dominance (SGLD), it consists of analysing whether (i) the most needy group of distribution x_2 generalised Lorenz dominates the most needy group of distribution x'_2, (ii) the two most needy groups of distribution x_2 generalised Lorenz dominates the two most needy groups of distribution x'_2, and so on with cumulated three most needy groups until the least needy group.

We shall consider attribute x_1 as a qualitative variable composed of $k = 1, ..., K$ groups ordered from the most needy to the least needy. We shall denote distribution of attribute x_2 related to the first k groups by $x_2(k)$.

> Definition 8: *Dominance in the Sense of the Sequential generalised Lorenz*
> *Given two bivariate distributions, distribution x_2 dominates in the sense of the sequential generalised Lorenz distribution x'_2, if and only if each distribution $x_2(k)$ dominates in the sense of Lorenz each distribution $x'_2(k)$, $\forall k = 1, ..., K$.*

Atkinson and Bourguignon (1987) consider a similar marginal distribution of needs in both distributions. Nevertheless, the sequential generalised Lorenz dominance criterion has been extended to the case where distributions of needs differ (Jenkins & Lambert, 1993; Chambaz & Maurin, 1998). Furthermore, Ooghe (2007) shows that sequential generalised Lorenz dominance can give different degrees of priority to different types of needs by giving higher weights assigned to more needy groups.

The needs approach is appropriate when we want to evaluate the consequences of changes in one attribute but fails if we are interested in the impact on the two attributes. Furthermore, if we consider that among the attributes one can be used to make direct transfers between individuals, then a compensation approach can be proposed (Muller and Trannoy, 2003). It is widespread in the literature on distributive justice to separate individual characteristics between those which are due to responsibility and those which are not. Opinions differ on which individuals characteristics are the concern of one or the other set of characteristics (Dworkin, 1981; Arneson, 1989; Sen, 1992; Roemer, 1998; Fleurbaey, 1995). Nevertheless, this literature provides ethical motivation for compensation perspective.

Assumption 2: One attribute can be used to compensate a disadvantage in the other and the compensation relies on a transfer sensitivity

Muller and Trannoy (2003) rely on a class of utilities functions which have ethical and intuitive meaning and they provide a test of welfare improvement in a multidimensional setting. Their main idea is that compensation is good for social welfare. In the income-health example, it means that transferring income from a healthy individual to a handicapped individual at given income is recommended. The compensation is particularly appropriate if handicapped people are in the bottom part of the income distribution. On the contrary, the healthier an individual is, the lower he claims for a redistribution, other things being equal. This recommendation imposes a negative cross derivative of the utility function between the compensating attribute and the compensated one.

The authors also introduce a transfer sensitivity according to which the policy makers are more sensible to transfer at the bottom of distribution than at the top. To do so, they assume a positive third partial derivative. This condition plays a significant role in a compensation approach where the two attributes are health and income. This transfer sensitivity can easily be illustrated in the French health system. If income is the compensated attribute, an example is the universal health care coverage so-called *Couverture Maladie Universelle*, which provides cover for people with lower income[13]: a wealthy individual does not seem to be a good candidate for social benefits even if he suffers from a poor health. Conversely, when health is the compensated attribute, an example is the

[13] This reform is more precisely described in chapter 4.

undertaking to reimburse medical expenses for individual who suffer from long standing illnesses[14]: a healthy individual does not seem to be a good candidate for social benefits even if he is poor.

Muller and Trannoy (2003) consider two classes of increasing utility functions concave in each argument and continuously differentiable to the required degree. In the first case, they assume that x_1 is the compensating attribute and x_2 is the compensated attribute. In the second case, they assume the reverse situation, x_1 is the compensated attribute and x_2 is the compensating attribute.

$$\mathcal{U}^1 = \{U_1, U_2 \geq 0, U_{11}, U_{12}, U_{22} \leq 0, U_{112} \geq 0\}$$
$$\mathcal{U}^2 = \{U_1, U_2 \geq 0, U_{11}, U_{12}, U_{22} \leq 0, U_{221} \geq 0\}$$

If we compare these classes with stochastic dominance at first order as proposed by Atkinson and Bourguignon (1982), it seems that $\mathcal{U}^2$ is intermediate between $\mathcal{U}^-$ and $\mathcal{U}^{--}$. The difference between these classes comes from the asymmetrical treatment between the two variables. The third cross derivative $U_{112} \geq 0 \in \mathcal{U}^1$ introduces the transfers sensitivity as well as the compensation. Indeed, when considered with $U_{12} \leq 0 \in \mathcal{U}^1$, it emphasises that attribute 1 can compensate for deficiencies in attribute 2. The compensation is particularly required when the distribution in the compensating attribute is low. The class $\mathcal{U}^1$ (respectively $\mathcal{U}^2$) represents a point of view where the researcher is primarily interested in the distribution of income (respectively health) among the unhealthy (respectively the poor). The positive third partial derivatives exhibits this point of view and means that the decrease in marginal utility of income (respectively health) is smaller among the healthy (respectively rich) individual than among the unhealthy (respectively poor) individual.

Trannoy (2006) use tools introduced in Muller and Trannoy (2003) in an explicit income-health example. His paper permits checking dominance for the above classes. The statistical tests to be implemented rely on the definition of the generalised Lorenz dominance (Shorrocks, 1983) presented in Definition 7 and extended to a multidimensional framework.

Definition 9: *Generalised Lorenz Dominance in multidimensional framework*
Given any two bivariate distributions x and x', such as $x = x_{ij}$ and $x' = x'_{ij}$ are defined in $\mathbb{R}_+^{n2}$, distribution x dominates in the generalised Lorenz sense distribution x', written $x \geq_{\mathrm{GL}} x'$, if
$\frac{1}{n}\sum_{i=1}^{k} x_{ij} \geq \frac{1}{n}\sum_{i=1}^{k} x'_{ij}$,
$\forall k = 1, ..., n$, $\forall j = 1, 2$.

Poverty gap dominance is known to be equivalent to stochastic dominance at second order in a univariate framework (Foster & Shorrocks, 1988), Trannoy (2006) adopts an extension of the absolute poverty gap for a poverty limit of x_1 (respectively x_2) in a multidimensional

[14] This is the so-called *prise en charge à 100% pour affectations de longue durée.*

framework as follows.

$$P_x(x_1|x2) = \frac{1}{n} \sum_{[i|x_{i1},x_{i2} \leq (x_1,x2)]} (x_1 - x_{i1})$$
$$P_x(x_2|x1) = \frac{1}{n} \sum_{[i|x_{i1},x_{i2} \leq (x_1,x2)]} (x_2 - x_{i2})$$

The associated dominance criterion is exhibited hereafter.

Definition 10: *Poverty Gap Dominance*
Given any two bivariate health distributions x and x',
distribution x health poverty dominates distribution x', written $x \geq_{P_1} x'$, if
$P_x(x_1 \mid x_2) \leq P_{x'}(x_1 \mid x_2)$, $\forall x_1, x_2$
distribution x income poverty dominates distribution x', written $x \geq_{P_2} x'$, if
$P_x(x_2 \mid x_1) \leq P_{x'}(x_2 \mid x_1)$, $\forall x_1, x_2$

Dominance criteria according to the asymmetrical classes can thus be obtained by the following sufficient conditions.

Proposition 1: *Dominance for classes $\mathcal{U}^1$ and $\mathcal{U}^2$*
Given any two bivariate health distributions x and x',

1. *if $x \geq_{\mathrm{GL}} x'$ and $x \geq_{P_1} x'$*
   then distribution x dominates distribution x' for the class $\mathcal{U}^1$, written $x \geq_{\mathcal{U}^1} x'$
2. *if $x \geq_{\mathrm{GL}} x'$ and $x \geq_{P_2} x'$*
   then distribution x dominates distribution x' for the class $\mathcal{U}^2$, written $x \geq_{\mathcal{U}^2} x'$

The two criteria require that the marginal distributions of the two attributes must be more egalitarian in the generalised Lorenz sense[15]. The condition related to poverty gaps is different according to the criterion. For the first criterion, it says that the health poverty gap does not increase for any levels of income and health. For the second criterion, it says that the income poverty gap does not increase for any levels of income and health. Moreover, this condition makes easy the use of these criteria in a poverty analysis by defining poverty thresholds for income and for health. The asymmetric perspective proposed by Muller and Trannoy (2003) and Trannoy (2006) offers to extend the notion of poverty gaps as well as the principles of transfers to the multidimensional context. The overview has also emphasised through the bidimensional illustration that multidimensional dominance fails in proposing an equivalent of the Hardy-Littlewood-Polya theorem.

[15]The two conditions stated in proposition 8 are sufficient. If we restrict the class of utility function to be transfer sensitive in income $U_{111} > 0$ and in health $U_{112} > 0$ then the conditions become necessary.

3.2.4 Orderings and rankings: some elements of conclusion

The use of ranking criteria has a direct intuitive appeal and as shown in this first section, it is also possible to give rigorous theoretical support to these intuitive approaches within the concept of the social welfare functions. This section has emphasised the availability of partial orderings for distribution. There are reasons to believe that the idea of inequality as a ranking relation may be inherently incomplete. Consequently, Sen (1973) stresses

> "a measure of inequality that involves a complete ordering may produce artificial problems, because a measure can hardly be more precise the concept it represents". In this context, it is worth saying that the need is for a measure that comes into its own with sharp contrasts, even though it may not provide a scale sensitive enough to order finely distinguished distributions.

The measures of inequality that have been proposed in the economic literature lie within a two categories framework (Sen, 1973). This first section has considered tools that attempt to measure inequality in terms of normative notions of social welfare. On the other hand, there are measures that try to catch the extent of inequality in objective sense. As regard to policy decisions against inequalities in health, if we want to exceed the simple declaration of intent, we need quantifiable goals. We move now to the evaluation of differences between distributions. There are also a number of measures of inequality that have been proposed in the literature. The following sections firstly exhibit those related to a unidimensional approach and then presents those used in a multidimensional approach.

3.3 Measurement of inequalities in health in a unidimensional context: the health Gini index

The first measurement tools of inequality in health were proposed in the Black Report (Townsend & Davidson, 1982). Two literature references are well-known in the field of the empirical measurement of inequalities in health, namely Wagstaff *et al.* (1991) and Kunst and Mackenbach (1996). Analogously to dominance criteria, indices of inequality in health have firstly been proposed in a unidimensional context and have then been extended to a bidimensional context[16]. The following section presents a tool for evaluating inequality in health in a unidimensional environment. Whereas, the approach proposed by Allison and Foster (2004) considers health as a qualitative variable, widespread traditional

[16]Beyond the fact that the bidimensional approach is technically an extension of the unidimensional approach, there is in the literature on inequalities in health a normative conflict on the approach to choose. The unidimensional approach is particularly recommended by the World Health Organisation (WHO, 2000), which argue for the widest measure of inequalities. Whereas the bidimensional approach puts forward the ability to represent precisely inequalities, which are also considered as the most inequitable inequalities (Braveman *et al.*, 2000).

instruments of distributional analysis such as Lorenz curve cannot be used with such qualitative or categorical indicators to evaluate inequality. Nevertheless, we assume that a health indicator is suitable for distributional analysis tools and present the traditional Gini index applied to health.

As presented in the previous section, the Lorenz curve was introduced in order to represent inequality using income share without considering differences in mean incomes. The Gini index is the inequality index associated to the Lorenz curve. It is a popular measure in economic research. Its application to health was firstly proposed by Le Grand (1987) in order to provide international comparative studies of health. The health Gini index is based on a Lorenz curve applied to health. This Lorenz curve plots the cumulative proportion of a population by increasing health status against the cumulative proportions of health. The Gini index equals twice the area comprised between the diagonal and the Lorenz curve. Its value is comprised between 0 when there is an equal distribution of health within the population, and 1 when all the population's health is concentrated in one person. In this latter extreme situation, the Lorenz curve is ⌋-shaped.

Formally, Brown (1994) provides a formula to calculate the Gini index in a population of n individuals, which is written

$$Gini = 1 - \sum_{i=0}^{n-1} (Y_{i+1} + Y_i)(X_{i+1} - X_i) \tag{3.6}$$

where X_i is the cumulated proportion of individuals ranked according to the outcome variable up to individual i, and Y_i is the cumulated proportion of the outcome variable up to individual i. Applied to health, the health Gini index relies on a ranking of individuals according to their health.

From a technical point of view, the Gini index has the advantage of being easily interpreted and many empirical works on inequalities in income as well as in health have used it. Gwatkin *et al.* (2000) even stress that the Gini index remains the most frequently used indicator as there is no clear consensus about a preferred alternative. The health Gini index is particularly useful for comparative studies either within various countries, region or across time. Among many other comparative studies based on Gini indices, we can quote Le Grand (1987), Brown (1994), Deaton (2001), van Doorslaer & Jones (2003), Jones & Lopez (2004), etc. Moreover, it should be noted that the measures of inequality in health proposed by the World Health Organisation (2003) as part of the ranking of health systems rely on Gini-based measures of inequality.

Despite its widespread use, this tool presents some disadvantages. Firstly, according to the Gini index, the desirable goal of equality is a uniform distribution of health. Nevertheless, this objective is clearly hard to reach. Secondly, two populations can present exactly the same value of Gini index and meanwhile distributions of heath in each population are very different. Indeed, the Lorenz curve can have different shapes but yields the same value of Gini index. Thirdly, the Gini index fails in differentiating specific situations. For example, a Gini index concludes a decrease of inequality in health in the case which a sick individual stays sick and a healthy individual becomes ill. Moreover, the Gini index measures inequality irrespective of the socioeconomic status of the persons concerned. Inequality measured by the Gini coefficient would also fall even if the healthy person getting sicker is poor. Most authors therefore argue that measuring inequality in health across individuals without taking into account any social dimension is not interesting (Wilkinson, 1997).

3.4 Measurement of inequalities in health in a bidimensional context

The bidimensional context permits taking into account the correlation of health with other indicators, especially income. According to the principles of action argued by the WHO (Whitehead, 1992), most of the inequities in health are due to living and working conditions. In that way, considering bidimensional inequalities permits to ask equity questions. In the ethical debate as well as in the policymakers' view, there seems to be agreement that some inequalities are inequitable and unjust, whereas some inequalities are unavoidable or legitimate[17].

The indicators for the measurement of bidimensional inequality of health could be divided into two groups, those which rank individuals according to the second attribute, e.g into a social hierarchy when income is the second dimension, and those which do not. The pseudo-Gini of health is the concern of the latter whereas the concentration index requires a rank.

3.4.1 The pseudo-Gini

The health Gini index measures inequality in health in the strict sense. Indeed, it does not distinguish situations where the sickest are also the poorest or the other way round. The same value of the Gini index could thus be calculated regardless of the form of distribution of health according to a particular dimension. In order to consider a social dimension in this tool, the Lorenz curve of health has been used differently from the way in which it was initially proposed by Le Grand (1987). Preston *et al.* (1981) propose a pseudo-Lorenz curve which plots the cumulative proportions of social groups ranked from

[17]This remark allows us to give importance to the use of inequality indices in bidimensional context. Nevertheless, we would not like to take part to the debate on the distinction between fair and unfair inequalities in health. We refer to the literature on this question (Fleurbaey, 2005; Fleurbaey & Schokkaert, 2007).

the sickest to the healthiest against the cumulative proportions of health. Analogously to the previous Gini index, a pseudo-Gini index is defined as twice the area between the Lorenz curve and the diagonal. Concretely, this measurement tool can be applied using a categorical social variable, which is ranked or not. Few studies on inequality in health have used the pseudo-Gini index (Preston *et al.*, 1981; Ruger & Kim, 2006). In addition to being hard to interpret intuitively, the pseudo-Gini index presents a major issue. As it ranks social groups according to the health status, it does not adequately take into account the hierarchical nature of the socioeconomic status variables (Kunst & Mackenbach, 1996). Therefore, the index does not differentiate between a situation where the group of the sickest is composed of socially-advantaged people and a situation where it is socially-disadvantaged people. Finally, the strong relationship that links health and social characteristics is ignored.

3.4.2 The concentration index

Proposed by Wagstaff *et al.* (1991), the concentration index is a measurement tool, which has been successfully used in the social inequality literature. It has been used it to compare the levels of inequalities in health in the European countries and to establish the role of economic factors in accounting for cross-country differences in inequality in health and intra-country health variations (van Doorslaer & Jones, 2003). The concentration index construction is close to the construction of the Gini index. This index relies on a health concentration curve, which plots the cumulative proportions of the population ranked by increasing social status against the cumulative proportion of health status.

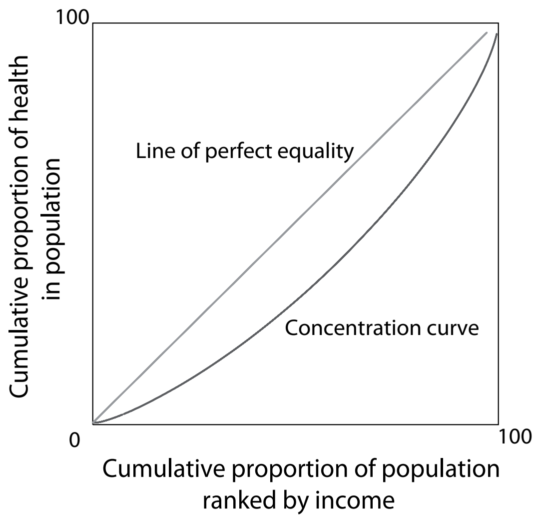

Figure 3.3: The health concentration curve

If health is equally distributed over the social dimension, the concentration curve coincides with the diagonal. If the sickest are concentrated among the most socially-disadvantaged, then the concentration curve is below the diagonal. The farther the concentration curve is from the diagonal, the higher the degree of inequality is. Conversely, if the concentration curve is above the diagonal, then the sickest are concentrated among the most socially-advantaged. The figure 3.3 gives a graphical representation of a concentration curve.

The concentration index is analogous to the Gini index and equals twice the area between the concentration curve and the diagonal. The concentration index takes values comprised between -1 and +1. It is positive (respectively negative) if the concentration curve is above (respectively below) the diagonal. The lowest (respectively the highest) value corresponds to the situation where all the health status in the population is concentrated on the most disadvantaged group (respectively the most advantaged group).

The concentration index is defined by the following formula

$$CI = 1 - 2\int_0^1 C(p).dp \qquad (3.7)$$

where p is the cumulative proportion of people ranked by increasing social information and $C(p)$ is the health concentration curve. If individuals are ranked in the same way according to health and the social variable, then the concentration index equals the Gini index.

The concentration index has some advantages common to those of the Gini index. Firstly, it provides both a quantified and a graphical measure of inequality. Secondly, it provides an inequality measure whose sign depends on the socioeconomic gradient of the inequalities in health. Its measurement is sensitive to changes in distribution of the population across the socioeconomic dimension (Kakwani *et al.*, 1997). Thirdly, it can be used to compare health distributions across different population conditional that the indicator of health is similar from country to country. Similarly, if the indicator of health is comparable from one period to another, this tool gives an indication of changes in inequalities over a period of time. Finally, unlike previous tools, the concentration index can be applied on both grouped and individual indicators. Therefore, the social indicator required in the concentration index can be both ordered categorical or continuous.

Moreover, the concentration index has some other properties that turn it into one of the most used measurement tool in studies of social inequality in health. We shall describe these four properties below.

1. **Confidence intervals**

   Although Mackenbach and Kunst (1997) underline that the relation between the magnitude of inequality and the value of the concentration index can be hard to interpret, a standard error for concentration index can be computed in the same way these authors do for the relative index of inequality (Kakwani *et al.*, 1997). The computation of standard error on individual data relies on a convenient ordinary least square regression as described in the following equation

$$\frac{2\sigma_r^2}{\overline{y}} y_i = a + br_i + \epsilon_i \tag{3.8}$$

Formally, $\overline{y}$ is the mean health status, σ_r^2 is the variance of the rank of the socioeconomic dimension and b is an estimate of the concentration index, which is equal to

$$b = \frac{2}{n.\overline{y}} \sum y_i r_i - 1 - \frac{1}{n} = CI \tag{3.9}$$

The standard error of b provides a standard error of CI. In the case where the assumptions for an OLS regression cannot apply, a Newey-West regression (Newey & West, 1987) is carried out and it corrects for both autocorrelation and heteroscedasticity observed in data.

2. **Decomposition of the concentration index**

One of the popular features of concentration index as a measure of bidimensional inequality in health is its ability to incorporate an econometric model for health with several control variables and subsequently proceed to the decomposition of inequality into the contribution of these regressors (Wagstaff & van Doorslaer, 2003). The decomposition relies on the assumption according to which the explained variable (i.e. health) is additive in its regressors. The decomposition method permits observing inequality and to identify its sources.

We assume that the following linear regression model defines the health status of individual i according to k regressors, such as $k = (1, ..., K)$

$$y_i = a + \sum_{k=1}^{K} b_k x_{ki} + \epsilon_i \tag{3.10}$$

The random error term, ϵ_i is assumed to have expected mean value equal to zero and constant variance. The b_k are assumed constant for every individual i. By substituting this equation in the concentration index formula (3.7), we obtain

$$CI = \sum \left(b_k \frac{\overline{x_k}}{\overline{y}} \right) CI_k + \frac{2}{\overline{y}} cov(\epsilon_i, r_i) \tag{3.11}$$

The concentration index is assumed to be made up of two components: an explained one, equal to a weighted sum of the concentration indices of the k regressors, and a residual component. The weight represents the estimated health elasticity with respect to regressor k, as

$$\eta_k = b_k \frac{\overline{x_k}}{\overline{y}} \tag{3.12}$$

The estimated inequality in heath is thus expressed as a sum of inequality in each of its determinants, weighted by their own elasticity with health. This decomposition method emphasises the contribution of each regressor to the explanation of the inequality. It gives each regressor's respective impact on health as well as the degree of inequality of this regressor's distribution with respect to the social dimension.

Therefore, various regressors can be considered: traditional socioeconomic ones, such as income, education levels, and activity status; geographical conditions, such as regions, areas or urban conditions; health insurance conditions; demographics as well as health status characteristics, such as limitations or healthy behaviours etc. The decomposition has the main advantage to permit a computation of the contributions of particular conditions on which policymakers may concentrate and intervene.

3. **Degree of inequality aversion into the concentration index**

When one is interested in inequality, it is usual to consider an inequality aversion term. Indeed, an individual would always prefer a society with a more equal distribution of income. The concentration index has an extended version that contains ethical judgements about inequality aversion. Analogously to the extended Gini proposed by Yitzhaki (2003), the extended concentration index uses a parameter to emphasise various parts of distribution of the social dimension, namely an inequality aversion term. As the concentration index represents an area, it can be expressed using an integral. Assuming that v is this inequality aversion parameter, the extended concentration index is written

$$CI(v) = 1 - v(v-1)\int_0^1 (1-p)^{v-2} C(p) dp \tag{3.13}$$

Therefore, $C(p)$ is defined for $v > 1$. The standard concentration index is given when $v = 2$. The higher this parameter v is, the more emphasis is given on the situation of the poorest individuals.

This extended index can be written in different ways in order to emphasise the subjacent ethical judgements. Firstly, Lerman and Yitzhaki (1989) provide a covariance-formula for the extended Gini index that can be applied to concentration such as

$$CI(v) = -\frac{v \times cov\left(y, (1-r)^{v-1}\right)}{\overline{y}} \tag{3.14}$$

Secondly, Wagstaff (2002) proposes to go further by using a formula of the concentration index emphasising a particular view where the income distribution reductions in inequality in health matter most. This extended concentration index is written as follows

$$CI(v) = 1 - \frac{v}{n\bar{y}}\sum_{i=1}^{n} y_i(1-r_i)^{v-1} = 1 - \sum_{i=1}^{n}\frac{y_i}{n\bar{y}}w_i \tag{3.15}$$

where $w_i = (1-r_i)^{v-1}$ represents the weight allocated to the share of health of individual i in the population. When $v = 1$ then $w_i = 1$ and health status of each individual in the population is equally weighted. Therefore, there is no aversion to inequality whatever the distribution of health over the social dimension. Wagstaff tests different values for v comprised between 1 and 8. He concludes that

> "as v is raised above 1, the weight attached to the health of a very poor person rises, while the weight attached to the health of people who are above the 55^{th} percentile decreases. (..) for $v = 6$ the weight attached to the health of persons in the top two quintiles is virtually 0. When v raised to 8, the weight attached to the health of those in the top half of the income distribution is virtually 0".

To conclude, the standard concentration index takes into account the aversion to inequality as $v = 2$ and favours individuals who are at the bottom of the distribution of the social dimension.

4. **Interpretation of the concentration index in terms of redistribution policies**

As the concentration index has a non-intuitive scale and fails in a straightforward interpretation in natural units, Koolman and van Doorslaer (2004) have proposed an interpretation of this index using two more intuitive measures of redistribution and correlation from income literature. Considering similarities that link the Gini and the concentration indices, the two authors uphold that the intuition from the Robin Hood index[18] and the Blackburn redistribution[19] can be applied to a health perspective by moving it from the measurement of inequality in income when ranked by income towards the measurement of inequality in health when ranked by income. Finally, the authors propose innovative interpretations of the concentration index and their results emphasise

> "a CI value that can easily be translated into a percentage redistribution required from rich to poor to make estimated income-related inequality equal to zero but not, however, to obtain equality".

[18] This index stems from the income literature and has both graphical and quantified versions (Kawachi & Kennedy, 1997). It is equivalent to the maximum vertical distance between the Lorenz curve and the line of equal incomes.

[19] This redistribution relies on a transfer of a fixed amount of the mean income level, also called a fixed lump sum amount, from all units above the median income to those below.

The concentration index is popular in the analysis of inequalities in health, particularly because it has four main advantages that have previously been discussed. The most important advantage is the ease with which it can be decomposed. Despite this popularity and beyond the limits that this index shares with the Gini index, several recent papers have emphasised its limits (Clarke *et al.*, 2002; Erreygers, 2006; Fleurbaey, 2005; Wagstaff, 2005; Fleurbaey & Schokkaert, 2007). Indeed, when the concentration curve crosses the diagonal, it indicates an inequality favouring the rich in some parts of the health distribution and favouring the poor in others. The areas above and below the curve can compensate each other (Humphries & van Doorslaer, 2000). Three other limits are relevant and described below.

1. **Can any health variable be used with a concentration index?**
   When computed from a binary health variable, the bounds of the value of the concentration index depend of the mean of this health variable (Wagstaff, 2005). The concentration index is supposed to take values between -1 and +1, but the bounds turn out to be much wider for population with a low mean, close to 0 than for populations with a high mean, close to 1. Therefore, as the mean increases, the range of the possible values shrinks, and the concentration index tends to equal 0 when the mean tends to equal 1.

   Similarly, the use of categorical variables is tricky. The concentration index for a categorical variable is hard to interpret unless categories are equidistant from one to another. However, in general, health variables do not have equidistant categories[20]. As the health concentration index is also a mean-based measure, this aspect can be related to Allison and Foster (2004) who criticise the applicability of these indices with such qualitative data.

   Therefore, when we apply the decomposition method of the concentration index, the use of a continuous health variable is preferred. In particular, such a variable allows researchers to rank precisely individuals. Nevertheless, as regard to the two first chapters of this dissertation, the availability of such health indicators is not straightforward.

2. **What about individuals' preferences?**
   Fleurbaey (2005) underlines the absence of individual's preferences in the concentration index. The traditional principle in inequalities in health care utilisation studies is to rely on the need principle. However, one can advocate that need characteristics are not only based on age, gender or objective health status but also rely on individual preferences or simply health expectations. Fleurbaey (2005) quotes the example

[20] For example, considering self-assessed health, someone reporting a very good health status does not have a health status two times higher than someone reporting fair health status.

of an individual whose professional career involves a lot of physical performance, and so this person is in greater need of a healthy body. His argument can be transferred to the question of income-related inequalities in health as inequalities in health status are traditionally standardised on the so-called unavoidable inequality (van Doorslaer & Koolman, 2004 ; Leu & Schellhorn, 2004) or policy irrelevant variables (Gravelle, 2003). This unavoidable inequality concerns inequalities due to biological differences like age and gender, which are to a large degree unalterable (Kakwani *et al.*, 1997). There is a tacit agreement on the vector on which inequalities in health are standardised that could be called into question. Nevertheless, it is clear that people have different expectations of health status (Moesgaard *et al.*, 2002). This question begets another question on the goal pursued by the concentration index.

3. **A concentration index equals to zero: a desirable goal?**
The standardisation of the measurement of inequalities in health on age and sex implies that the inequality of interest is the inequality due to social characteristics. The underlying policy goal is thus to neutralise these inequalities. In other terms, the objective is to nullify the correlation between health and socioeconomic conditions. This objective gives rise to various criticisms. Firstly, we need ask if the optimal degree of correlation between health and socioeconomic status is zero. Indeed, this condition ignores the existence of a compensation link between health and income, according to which a greater health can compensate for a lower income. Secondly, this condition goes against individual preferences as presented in the previous item, because individuals could consider health as any other good and decide whether they want to give priority to this good or to another one.

   Finally, Fleurbaey (2007) argues that reducing this correlation is not a defensible objective for health public policies. He rather proposes to improve the situation of the worst-off in terms of healthy-equivalent income, which is an index of living standard taking into account individual preferences about the relative importance of health among other dimensions.

3.5 Conclusion

The measures of inequality that have been proposed in the economic literature lie within a two categories framework. Sen (1973) distinguishes, on one hand, the analysis of inequality in terms of normative notion of social welfare, and on the other hand, the analysis of inequality in the objective sense. The choice between which of the two approaches to pursue is not easy and is even undesirable as when measuring inequality we need both facts and opinions.

Our presentation has introduced a second level of division within this two categories framework: the number of attributes to consider. The unidimensional approach concerns a lonely attribute whereas the multidimensional approach considers more than one attribute and contemplate the symmetrical or asymmetrical aspects of these attributes. From a general point of view, it seems that in a multidimensional context, attributes are rather substitute than complement in individual well-being. Therefore, a higher endowment in one attribute could compensate for a lower dotation in the other. This conclusion is true for the dominance as well as for bidimensional indices, such as the concentration index. The use of a compensation relationship between the attributes of individual well-being offers relevant perspectives for the measurement of inequality with an interest for subjacent public policies acting on income to improve health.

To conclude, we stress that what matters in measuring inequalities in health is to be able to conclude unambiguously on their existence and to evaluate its scale. Another important element is to go further than a pure measurement of inequality in health and to use additional information, such as other attributes (e.g income, education, gender), needs, individual preferences, normative aspects, etc.

Chapter 4

Income-related inequalities in health in France

4.1 Introduction

In France, studies on disparities in mortality, specific health problems and disabilities caused by socioeconomic status are particularly well-documented (Leclerc *et al.*, 2000; Jusot, 2003; Boissonnat & Mormiche, 2007). They emphasise very large inequalities in mortality in France. Lower socioeconomic groups are known to have higher rates of mortality than higher socioeconomic groups (Girard *et al.*, 2000). Moreover, inequalities in health between social groups seem to have increased over time and would be higher in France than in other European countries (Kunst *et al.*, 2000). While over the period 1976-1984 the mortality rate of French blue collar workers aged 35-80 years old was 1.8 times higher than the mortality rate of their white collar counterparts, the ratio increased to 1.9 between 1983 and 1991 and reached 2.1 between 1991 and 1999 (Monteil & Robert-Bobée, 2005). Nevertheless, few French studies concern inequalities in health as measured by more global health indicators (Chauvin & Lebas, 2007). Lack of this became particularly noticeable when income-related inequalities in health have been widely explained in Europe using more global health indicators: e.g Gravelle and Sutton (2003) in Great Britain, Lecluyse (2007) in Belgium, Leu and Schellhorn (2006) in Switzerland or Garcia and Lopez (2007) in Spain.

One of the challenges in measuring inequalities in health is to have a usable health measurement. Besides mortality or life expectancy, health status does not have a cardinal nature. In this context, the field of the measurement of health status has had an increasing interest, with recent propositions for sophisticated channels to transform an ordinal health measure into a continuous variable. To our knowledge, this sophisticated technique has not been applied to French data[1]. In this chapter, we carry out an analysis of inequalities in health with different measurements of health. Firstly, we replicate the new approach of measurement proposed by van Doorslaer and Jones (2003), which cardinalises self-assessed health using estimated thresholds from the Canadian Health Utility Index. Besides the HUI questionnaire not being available in France, the universality of this index can be called into question, and so we consider alternative measurements. The second measurement of health is an adaptation of the previous approach but relies on a generic distribution of health in the French population. Then, we consider the innovative continuous health index generated in the chapter 2 as a third measurement of health and test its reliability in an empirical study. Moreover, because the French self-assessed health is reported on two different types of scale of responses in our data[2], we also consider the effect of this feature on the subjacent analysis of inequalities. The three previous measurements of health are applied to self-assessed health reported over five response categories, and the last two measurements of health also consider self-assessed health reported in eleven categories. As a result, the second aim of this chapter is to give a comprehensive understanding on the measurement of health within the analysis of inequalities in health and, in particular, to show whether our health index is reliable for the measurement of inequalities in health.

Taking into consideration that the French government has introduced a new program in 2000, called *Couverture Maladie Universelle* (CMU), the study of inequalities in health is particularly relevant if it concerns two points in time. This program supplements the health insurance coverage provided by the Social Security system and reduces the out-of-pocket cost incurred by low-income patients. While the existence of financial barriers to care prior to the introduction of this program is undisputed, it is important to highlight the situation after its implementation. We thus compare inequalities in health in 2004 with inequalities in health in 1998. We focus particularly on characteristics which strongly contribute to socioeconomic inequalities in health in 2004 and analyse their contribution in 1998 as well as their changes over time. We are interested in understanding the impact of CMU on health and not on health care. Indeed, interest in equality in access to health care is derived from inequality in health as it may be one of the causes for inequality in health (Wagstaff & van Doorslaer, 2000a; WHO, 2000). The analysis of the effects of CMU on health status has not been carried out so far. Nevertheless, as we know that this free coverage has influenced health care utilisation in the meantime, we can expect to observe some effects on individuals' health too. After four years of implementation, it is unclear if it is still too early to assess relevant effects on the health status of the targeted individuals.

[1]The only study of income-related health inequalities which involves France is the comparative analysis carried out by Van Doorslaer and Koolman (2004) using the 1994 European Community Household Panel.

[2]Individuals are asked to report their health status in five categories from "very poor" to "very good" and they are also asked to evaluate their health status on a scale from 0 to 10.

The second section presents the French health care system and offers a primary analysis of disparities of health caused by socioeconomic status in 2004 using stochastic dominance tools. The third section describes the measurements of health which are involved in the analysis. The fourth section focuses on the measurement of income-related inequalities in health in 2004. The fifth section describes these inequalities in health in 2004 in detail by decomposing them into contributing factors. Furthermore, these two sections also offer a comparative analysis of inequalities in health according to the measurement of health, i.e. according to the mapping used to cardinalise self-assessed health. The sixth section considers income-related inequalities in health in 1998 and offers an analysis of changes in inequalities in health between 1998 and 2004. Conclusions are presented in the last section.

4.2 Health in France in 2004

4.2.1 The French health care system over the last decade

The French health care system is based on the principle of horizontal equity, according to which individuals with equal needs should have identical access to care regardless of their socioeconomic status. A series of changes in the French health care system over the last ten years have given rise to a new concern for inequalities in health. The great majority of the French population, namely 98%, is covered by the Social Security system. Nevertheless, the compulsory national health insurance fund only covers between 70% and 80% of total health care cost. Patients face user charges when they visit general practitioners as well as specialists when they stay at the hospital or buy drugs, optical or dental prostheses (Couffinhal & Paris, 2003). Therefore, individuals can purchase voluntarily a supplementary medical health insurance to cover these charges. These private insurance policies are usually funded through flat-rate premiums, which are sometimes subsidised by employers. The poorest individuals such as unemployed people with no social benefits or homeless people or other socially-disadvantaged people are less often covered by a private insurance. It should be noted that the poorest French population was still uncovered by a supplementary health insurance in 2000. Consequently, one of the most striking policy changes in that year has been the extension by law of free access to medical care to a larger number of individuals with low income through a universal health care coverage, called the *Couverture Maladie Universelle* (CMU). Besides granting access to compulsory medical insurance, this reform has provided the poorest 4.5 million individuals with a free supplementary health insurance and has also exempted them from out-of-pocket payments and

Avance de frais[3] for their health care consumption (Boisguérin, 2005). Almost 4 million people were automatically enrolled when the plan began in January 2000.

Prior to the implementation of CMU, a limited coverage was granted to the poorest and sickest through the *Aide Médicale Générale* (AMG) but this social benefit varied substantially across French departments. Its extent was, however, fairly limited, mainly exempting individuals from having to pay the *ticket modérateur*[4] while offering no cover for balance-billing by providers or for optical or dental care. At the time the CMU was introduced, the AMG accounted for about 3% of the population. These individuals were automatically enrolled in the CMU.

Since its introduction, the effects of the CMU on health care utilisation have been analysed. In particular, it has been shown that CMU beneficiaries use more health care *ceteris paribus* than any other people having a supplementary health insurance (Raynaud, 2003; Grignon & Perronnin, 2003). This impact on health care utilisation is explained by the poorer health status among socially disadvantaged groups. Nevertheless, this impact can also be explained by a moral hazard in the behaviour of CMU beneficiaries: those who enrolled on the plan may be those who expect to use health care more (Grignon *et al.*, 2007). Concerning inequity aspects, Huber (2006) shows that the introduction of the CMU explains most of the reduction of the horizontal inequity index of health consumption between 1998 and 2002. However, the efficacy of this programme in reducing social inequalities in health status has not yet been fully assessed. The only outcome measure is that by the end of 2000, CMU beneficiaries declared that their health status had improved during that year more frequently than non-beneficiaries (Raynaud, 2003).

As regard to the effects of CMU on inequalities in health care utilisation, we can intuitively foresee the effects on inequalities in health. The first element that we might observe is a selection bias of CMU beneficiaries according to which those who enrolled the plan may also be those who are in poorer health. The second element is that four years is a short time particularly when improvement or recovery of health requires a longer time period for a poorer health status. Nevertheless, our study provides a useful first step in the evaluation of this reform on inequalities in health, as in a long term the true goal of CMU is not only to give a chance to the poorest individuals to use health care but also to provide them with a better health status.

[3]In France, doctors' consultations as well as drugs must be paid for in full (with certain exceptions) at the time of use and reclaimed afterwards.

[4]The *ticket modérateur* is the co-payment, after refunding by the French National Health Insurance fund. It is a minimum contribution of the person insured to his/her health expenses. It can be partly or fully refunded by the supplementary individual health insurance. See Rochaix and Hartmann (2005) for an extensive presentation of recent changes in the French health care system.

4.2.2 Health in 2004: a social health gradient in France?

In this subsection, we would like to give an idea of the situation of inequalities in income and disparities in health in France in 2004. Although standard usage is to present methods before the data, we allow ourselves not to follow this practice and present some descriptive elements of our dataset instead.

The data

We rely on data coming from the 2004 IRDES-HHIS[5] (Health and Health Insurance Survey). Whereas it is widespread in the literature to study inequalities in health status on the population aged 16 and over (Gakidou *et al.*, 2000; Gravelle & Sutton, 2003; Humphries & van Doorslaer, 2000), we point out the relevance of analyses according to age groups in order to take into account changes in individuals' health preferences due to age. We restrict our analysis to the working-age population, i.e. individuals aged 16 to 65 years old. The particular relationship between health and ageing justifies this restricted sample. Indeed, health status worsens with age and so is less influenced by income after 65 years old old. For instance, needs for health care is shown as less unequal among elderly people than among young age classes (Huber, 2006). Similarly, there seems to be no or limited income-related inequality in ill-health among persons aged more than 65 years old and inequality differences are highly significant between persons aged more than 65 years and persons aged less than 65 years old. According to van Ourti (2003), another reason for this difference in inequalities in health according to age is the income concept. A ranking based on permanent income is different from a ranking based on current income and, as a result, it potentially leads to a different degree of socioeconomic inequalities in health. As many surveys on health, the IRDES-HHIS dataset does not give a very detailed income information. As a consequence, in order to distinguish between permanent and current income, we would have to rely on arbitrary assumptions. For these reasons, we restrict our study to individuals who are under 65 years old. Our analysis relies on $8,235$ individuals in 2004.

We use household income as the measurement of income. In IRDES-HHIS questionnaires, households are asked whether each of them has different income and other financial resources. If so, these incomes are either detailed or at least reported as a global amount. From these answers the amount of current total disposable monthly income (everything included) is generated by statisticians in IRDES. Furthermore, households have to point out a category for their income. In this manner, if households do not know their global or detailed income, they only give a category. This is the case for $18,56\%$ of the whole sample. We use this information to generate a continuous income. Indeed, we calculate the income

[5] For a detailed presentation of the dataset, please refer to chapter 2.

median per income category and replace unknown monthly income by the median. We then transform this income variable into real terms[6]. This income in inflation-adjusted euros is then transformed into a household income per consumption units using the modified OECD scale, which gives a weight of 1 to the first adult, 0.5 to the second and subsequent adults, and 0.3 to each dependent.

Inequalities in the distribution of income

Studies of inequalities in health tend to focus on relationships between socioeconomic factors and health. In particular, it is widely acknowledged that individual income is a powerful determinant of individual health (Wagstaff & van Doorslaer, 2000b; Subramanian & Kawachi, 2004). It is thus relevant to evaluate the inequality of income prior to an analysis of income-related health inequalities. Various methods are available to describe and quantify the extent of inequality of income within a given community or society. Of these, the Lorenz curve and the Gini coefficient are frequently used. As described in chapter 3, the Gini index is related to a Lorenz curve and equals twice the area comprised between the diagonal and the Lorenz curve. A graphical representation of the income Lorenz curve in the 2004 IRDES-HHIS is proposed in figure 4.2. The income Gini coefficient associated to the household income per consumption units available in our dataset equals $0,29$ as shown in table 4.1.

Variable	Gini Index	Newey-West S.E	[Conf. Interval] 95%
Household income per CU	0,2941	0,0073	[0,2798; 0,3084]

Figure 4.1: Income Gini index in 2004 (*2004 IRDES-HHIS*).

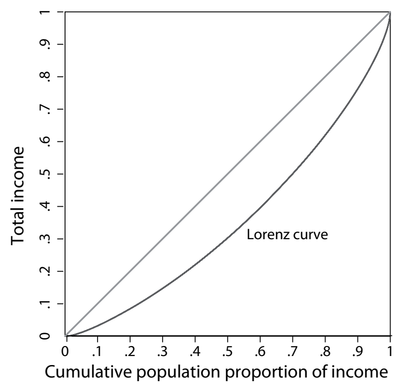

Figure 4.2: Lorenz curve for the household income per CU(*2004 IRDES-HHIS*).

A Gini coefficient equal to 0 represents a situation of perfect equality, which is not observed here. There is a clear inequality in the distribution of income. Nevertheless, when considered alone, the Gini coefficient has a limited interest, and so we shall compare this

[6]We use the 2004 consumption price index, with year 1998 as base. The consumption price indices are available from INSEE `http://www.insee.fr/fr/indicateur/indic_cons/historique.asp`.

value to other income Gini coefficient, such as the actual income Gini coefficient in France in 2004. According to Landais (2007), a Gini of 0, 44 is observed in 2004, which means that the income Gini in the 2004 IRDES-HHIS is lower than the actual Gini coefficient. Nevertheless, our income Gini coefficient relies on a restricted sample of individuals and a lower value is expected in this restricted context. Moreover, if we refer to the UNU-WIDER World Income Inequality Database (WIID), which collects and stores information on income inequality for developed, developing, and transition countries, the income Gini in France in 2004 equals 0, 28, which is much closer to the value we obtain. Secondly, we shall also compare the French Gini index to other Gini values for European countries. In the UNU-WIDER database, the income Gini coefficient is 0, 26 for Belgium, 0, 31 for Spain, 0, 32 for Germany and 0, 33 for Italy. The income inequality in France thus seems less important than it is in Italy even if these countries have close values (e.g, in Mexico, the income Gini coefficient is 0, 50 and in Sweden, it is 0, 23).

Distribution of health and income over the population: some health indicators

The IRDES-HHIS contains various health indicators. The table 4.1 presents some descriptive statistics of these health indicators over the sample.

Health variables	Proportion (%)	Mean of equivalised household income
Chronic disease		
No	83,98	1956,10
Yes	16,02	2142,14
Self-assessed health		
Very good	60,01	2011,31
Good	25,79	2039,79
Fair	13,32	1727,94
Poor	0,83	1504,09
Very poor	0,05	843,18
Disability level		
0 No discomfort	16,47	1753,64
1 Very weakly hampered	22,68	1993,61
2 Moderately hampered	35,41	2054,74
3 Hampered but normal life	21,76	2035,19
4 Limited professional/domestic activity	3,4	2078,37
5 Highly hampered	0,24	1345,92
6 No autonomy for domestic activities	0,04	1451,23
Vital risk		
0 No vital risk	38,91	1951,21
1 Prognosis very weakly pejorative	16,07	1994,77
2 Prognosis weakly pejorative	36,72	2004,89
3 Possible risk on vital conditions	7,93	2001,83
4 Prognosis probably bad	0,34	3130,83
5 Prognosis certainly bad	0,04	1840,78

Table 4.1: Descriptive statistics of some health indicators in the 2004 IRDES-HHIS.

Generally, respondents are in good or very good health: only 16% individuals suffer from at least one chronic disease; three quarters of the sample report a good or very good health[7]; one third of the sample is hampered but have a normal life whereas almost 40% have no discomfort or a weak disability. Concerning vital risk, a possible vital risk concerns almost 8% of the sample. The third column of table 4.8 gives the average household income by health indicators. Except self-assessed health which clearly presents an average income decreasing with a poorer health status, the other health indicators do not show a clear increasing or decreasing relationship with income. Nevertheless, people who are highly hampered by disability or have no autonomy for domestic activities have a lower average income. Analogously, individuals with a certainly bad prognosis of vital status have a lower average income than individuals with no vital risk. Chronic diseases seem to be less linked to income as the average income is not much different for people with chronic diseases and people without a chronic disease.

Inequalities in the distribution of health

It is worth to supplement these first descriptive statistics with some other statistics on a more global health indicator. We shall now consider the health index generated in chapter 2 and understand how health, as measured by the health index is unequally distributed over some individual characteristics such as age, income, education and economic status. Empirically, we rely on a graphical representation of cumulative distribution functions and on tests of stochastic dominance at first order as described in Lefranc *et al.* (2004) and in appendix A.

1. Distribution of health over age classes

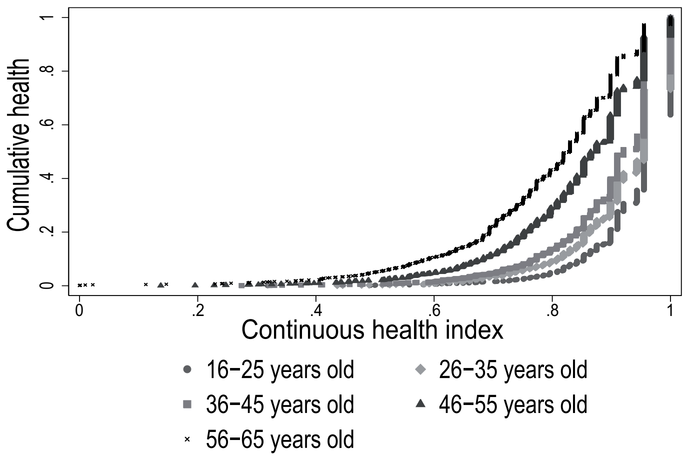

Figure 4.3: Empirical distribution function of the health index per age classes.

[7]We remind that the sample is reduced to individuals aged 16 to 65 years old.

As expected, figure 4.3 emphasises that health status worsens with age. We carry out dominance tests based on a conjunction of *Kolmogorov-Smirnov* unilateral tests to compare distributions of health over age classes. They confirm that people aged 16-25 years old are significantly in better health than all the other age classes (cf. table[8] 4.2) and that each age class is always dominating the upper age classes.

	16-25 y.o	26-35 y.o	36-45 y.o	46-55 y.o	56-65 y.o
16-25 y.o		<0,0001***	<0,0001***	<0,0001***	<0,0001***
26-35 y.o	0,999		<0,0001***	<0,0001***	<0,0001***
36-45 y.o	1	1		<0,0001***	<0,0001***
46-55 y.o	1	1	1		<0,0001***
56-65 y.o	1	1	1	1	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.2: P-value of *Kolmogorov-Smirnov* test related to health according to age classes

2. Distribution of health over income classes
   We consider now this health index according to income quintiles. The empirical distribution of the health index shifts to the right as income increases as described in figure 4.4.

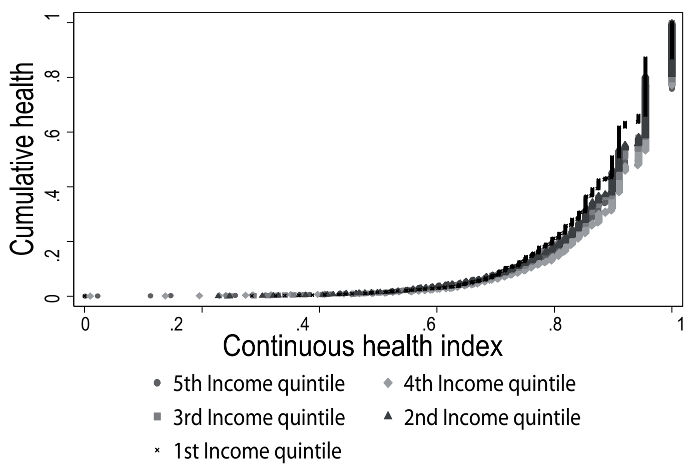

Figure 4.4: Empirical distribution function of the health index per income quintiles.

It emphasises that higher income levels experience a higher health status. Moreover, we carry out a dominance analysis and results emphasise that the two lowest income quintiles are significantly dominated at first order by the other three quintiles (cf. table 4.3). Individuals in the highest income quintile significantly have a better

[8]Explanation of the table: the result of the unilateral *Kolmogorov-Smirnov* test is read in row. The distribution of health of an individual aged 26-35 years old significantly dominates the distribution of health of individuals aged 36-45 years old, 46-55 years old, 56-65 years old as p-value<0,0001.

health status than individuals in any other income quintiles. The distribution of health is not significantly different between individuals being in the third or the fourth income quintile. Unexpectedly, the distribution of health of individuals in the first income quintile dominates the distribution of health of individuals in the second income quintile. The income-health gradient, as defined by Deaton (2002) is thus not observed in the two poorest income quintiles. Indeed, there is a gradual relationship between health and income according to which health improves with income throughout the income distribution, from middle income levels.

	1^{st} quintile	2^{nd} quintile	3^{rd} quintile	4^{th} quintile	5^{th} quintile
1^{st} quintile		0,74*	0,929	0,990	0,898
2^{nd} quintile	0,756		0,978	0,978	0,971
3^{rd} quintile	0,043**	0,001***		0,871	0,968
4^{th} quintile	<0,0001***	<0,0001***	0,207		0,829
5^{th} quintile	<0,0001***	<0,0001***	<0,0001***	<0,0001***	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.3: P-value of *Kolmogorov-Smirnov* test related to health according to income quintiles

3. Distribution of health over activity statuses

   The analysis of distributions of health according to activity statuses emphasise the two previous results : in figure 4.5, it is clear that “students” and “employed” experience a better health than “retired”, “inactives” or “homemakers”. In other words, younger age classes and higher income levels have a better health. We supplement the graphical analysis by unilateral tests whose P-values are presented in table 4.4. The distribution of health of “students” significantly dominates the distribution of health of all the other activity status. “Unemployed”, are significantly in worse health than “employed” people, which has already been shown in other empirical studies (Khlat & Sermet, 2004). The distribution of health of “retired” people is dominated by the distribution of health of all the other activity status, which is explained by the strong link between health and age. The distribution of health of “inactives” is only dominated by the distribution of health of “students”. Therefore, it seems that inactivity is not so much explained by a worse health status. On the contrary, the distribution of health of “homemakers” is significantly dominated by the distribution of health of “employed”, “students” and also “unemployed”.

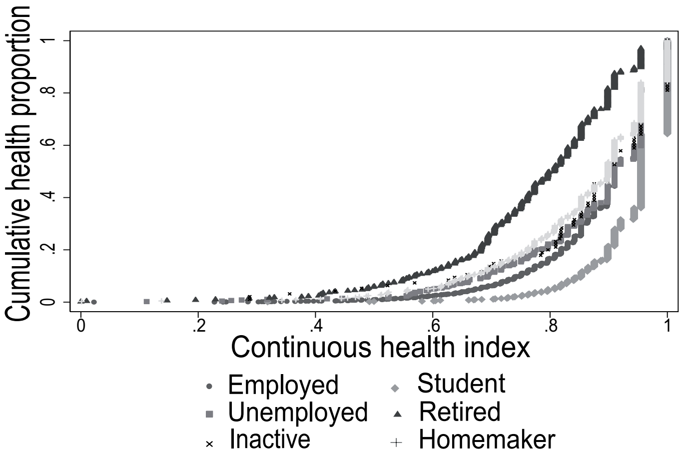

Figure 4.5: Empirical distribution function of the health index per socioeconomic statuses.

	Employed	Student	Unemployed	Retired	Inactivity	Homemakers
Employed		1	0,036**	<0,0001***	0,190	<0,0001***
Student	<0,0001***		<0,0001***	<0,0001***	<0,0001***	<0,0001***
Unemployed	0,992	1		<0,0001***	0,417	0,011**
Retired	1	1	1		0,974	1
Inactive	0,976	1	0,857	<0,0001***		0,257
Homemakers	1	1	0,984	<0,0001***	0,851	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.4: P-value of *Kolmogorov-Smirnov* test related to health according to socioeconomic statuses

4. Distribution of health over education levels

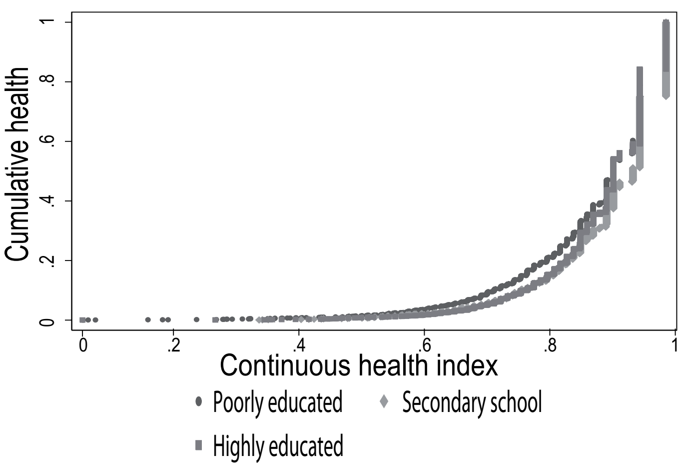

Figure 4.6: Empirical distribution function of the health index per education levels.

As for education level represented in figure 4.6, the distribution of health of poorly educated individuals (i.e those having no diploma) is situated on the left of the distributions of health of the two higher education levels. The unilateral tests emphasise that the distribution of health of individuals having at least A-level significantly dominates the distribution of health of individuals having either no diploma or a diploma of primary or secondary level.

	Education 3	Education 2	Education less
Education 3		0,231	<0,0001***
Education 2	0,839		<0,0001***
Education less	1	1	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.5: P-value of *Kolmogorov-Smirnov* test related to health according to education levels

This stochastic dominance analysis confirms the existence of social inequalities in health to income levels, activity status and education levels. Having considered these primary results, we are now interested in a parametric analysis of inequality which would consider health according to income and other individual characteristics.

4.3 Measuring inequalities in health: which measurement of health should be used?

Our analysis relies on a global measurement of health. We use individual self-assessed health. This variable has been comprehensively studied and criticised in chapter 1. The main disadvantage of this variable in the context of the measurement of inequalities in health is its ordinal categorical aspect. To analyse income-related inequalities in health, we need to cardinalise the information contained in self-assessed health. Several methods have been proposed in the literature. The more recent and more promising method is the method proposed by van Doorslaer and Jones (2003) which relies on a mapping from a generic health measure to the latent variable subjacent to self-assessed health. We propose five alternative mappings for self-assessed health based on this methodology. The first mapping is produced by applying the estimated thresholds of HUI to our self-assessed health variable as proposed in van Doorslaer and Jones (2003). Nevertheless, we put forward the reliability of the HUI thresholds for the French self-assessed health and look for a generic health measure available for the French population. The second indicator thus relies on self-assessed health cardinalised on SF6D. The third indicator is the health index generated in chapter 2. Considering that self-assessed health in 2004 IRDES-HHIS is available on two different scales, i.e a 5-points scale and an 11-points scale, the two latter indicators are also generated on the second scale.

4.3.1 New approach to measurement of health in Europe: an application to French data

Methodological strategy

Van Doorslaer and Jones (2003) propose to use the HUI predicted thresholds of each self-assessed health level to compute an interval regression on self-assessed health, even if the survey does not contain any generic health distribution similar to the HUI. Therefore, the same predicted thresholds have been used in some European studies and assumed that distributions of health in any European country were comparable to the Canadian distribution of health. We follow this suggestion and assume that there is a stable mapping from HUI that determines self-assessed health. This stable mapping applies not only to Canadian but also to French people. The actual thresholds are 0, 0.428, 0.756, 0.897, 0.947 and 1 for the best possible health status. In concrete terms, we compute these estimated thresholds in an interval regression model of the French self-assessed health in five categories. The interval regression model also includes different regressors. In this context, this is the level of HUI that is predicted considering that an individual has some particular characteristics Z_i. These characteristics give information on the individual's social conditions such as equivalent income, activity status, and education level.

Discussion

Although the Canadian distribution of health status is likely to be similar to the distribution of health in Europe and *a fortiori* in France, the authors' hypotheses need to be discussed from a general point of view and also in the French context.

Firstly, there are cultural differences in the way people report less than "good" health (Mackenbach, 2005). Using SHARE data, large differences in general indicators of physical health such as self-assessed health, long-standing health problems, and activity limitations have been emphasised between countries (Borsch-Supan *et al.*, 2005). For instance, when it comes to self-assessed health, German people are likely to rate their health more negatively than Dutch or Danish people, and the same applies to Italian and Spanish people as compared to French and Greek people. Furthermore, perceptions of "*excellent*" and "*very good*" health are varying with the cultural context and cannot be assumed to be identical, even in terms of frequencies. Indeed, when comparing the same sample of people answering both wordings[9], we clearly observe that the distribution of answer frequencies moves on the right when "excellent" is the highest category instead of "very good". In the French context, the spelling of self-assessed health in 2004 is different than the spelling used in the Canadian NPHS. In the Canadian version, self-assessed health is based on the simple question "*In general, how would you say your health is?*" and a choice among five possible answers: "*excellent, very good, good, fair and poor*". However, in the French questionnaire, for the analogous question: "*How is your general health status?*" the five proposed answers are "*very good, good, fair, poor and very poor*". In this context, the Canadian "*very good*" category corresponds to the French "*good*" category, the "*good*" to the " *fair*" and the " *fair*" to the "*bad*". Thus, the percentages[10] of each previous couple (Canadian self-assessed health category/French self-assessed health category) are very different: 37.1% versus 47.3% ; 27.1% versus 21.8% and 4% versus 8.6%. We believe that this dissimilarity between the two questionnaires could lead to a misleading measurement of health and therefore, to misleading results on inequalities in health.

Secondly, this mapping can only be applied to the 2004 IRDES-HHIS even if the IRDES dataset is available every other year since 1988. Indeed, the IRDES-HHIS survey from 1988 has collected self-assessed health on an 11-points scale (rate from 0 to 10) and the 5-categories question has been added to the usual questionnaire for the first time in 2004[11]. Historically, France has experienced an important debate on the way to ask self-assessed health in health questionnaires, especially on the number of response categories to propose. A scale in five categories was particularly criticised because individuals making a choice among an odd number of categories would be more likely to choose the medium category. We can thus quote several surveys carried out in the nineties where self-assessed health is asked over 2 to 11 categories: the 1997 INSEE survey "*Enquête permanente sur les conditions de vie*" (EPCV) collects a self-assessed health in six categories from 1=very high to 6=very weak; the survey "*Histoire de vie*" asks people whether they consider themselves to be in good health "*yes/no*", sick "*yes/no*", disabled "*yes/no*", old "*yes/no*"; the 1986 National Health Survey proposes five categories "*very good, good, fair, mediocre, frankly very poor*" and a sixth category for "*I do not know*". Finally the IRDES-HHIS questionnaire propose to respondents to evaluate their health status on a scale from 0 to 10 since 1988. This means that the implementation of the new approach of measurement on French health surveys is limited. If, for example, we want to analyse inequalities in health

[9] In SHARE 2004, both wordings are included in the questionnaire. As chapter 5 uses these data, differences between the two wordings were analysed. The following simple reproduction of the frequencies of both wordings confirms our comment. Wording 1 spreads out over *very good*, 13.8%; *good*, 48.2%; *fair*, 28.5%; *bad*, 7.3%; *very bad*, 2.2% whereas wording 2 is *excellent*, 7.8%; *very good* 14.9%; *good*, 43.9%; *fair*, 24.35%; *bad*, 9%.

[10] From the sake of comparability, the percentages concern a sample of individuals aged 16 and more in the 2004 IRDES-HHIS.

[11] The 2002 IRDES-HHIS has also collected a 5-points self-assessed health but only for head of household.

with the IRDES-HHIS in years prior to 2004, then we cannot used estimated thresholds as we do with the 5-categories self-assessed health in 2004.

We conclude that the distribution of HUI seems to be valid at a given time, for a given population, in a given context.

Nevertheless, the interval regression presents several advantages. Firstly, this method avoids the inappropriate use of ordinary least squares (OLS) to model an ordinal categorical variable. Secondly, it considers a vector of individual characteristics which leads to greater individual-level variations in the measurement of health. Finally, the interval regression considers external individual information to scale the categories of self-assessed health, which outperforms a construction based on arbitrary rescaling that could predict health status values out of the $[0;1]$ interval. Indeed, if a health distribution such as HUI is available for the sample, then the range of average values of this distribution for various age groups could be used. The same model is thus carried out with the distribution as the explained variable. The minimum and maximum predictions from this new model then define the observable range of the distribution conditional on the set of regressors. A similar extensive comparison of cardinalisation methods has been conducted using the 15D score from a Finnish sample. This study confirms that the interval regression is superior to ordinary least squares and ordered Probit (Lauridsen *et al.*, 2004). It is thus advisable to use a health distribution coming from the same context of the ordinal categorical health variable like it has been done in Belgium. They use the same method but scale self-assessed health on another continuous health measure, namely EQ-5D (Lecluyse & Cleemput, 2006).

As a result, there is a great interest in finding a generic health measure analogous to the HUI available for France: it will allow us to use an innovative cardinalisation method. Moreover, as the 2004 IRDES-HHIS allows the use of the HUI thresholds with a 5-categories self-assessed health, then we will be able to compare alternative mappings.

4.3.2 Cardinalisation of self-assessed health: a reliable health distribution in France?

In France, SF36 is the only generic health measure with an empirical distribution which is available at a general population level. It is included in the 2003 French National Health Survey (Leplège *et al.*, 1998; Leplège *et al.*, 2001).

Is SF36 analogous to HUI?

The SF36 Health Survey is a standardised questionnaire used to assess patient health across eight dimensions of health[12]. It consists of items or questions on each health di-

[12]SF36 yields an 8-scales profile of functional health and well-being scores, namely physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health.

mension. Responses to the items are combined into dimension scores. These scores permit describing health differences between patient groups or from one time period to another. Despite the fact that dimension scores range from 0 to 100 (the higher health-related quality of life), they are not comparable and there is no basis for combining them into a single index. In particular, scores rely on a simple arithmetic aggregation.

However, the interval regression method relies on a generic health distribution being cardinal. SF36 is a health profile measure. This type of measure is known to have the weakness of not always allowing judgements of which of the two profiles is better than the other (Nord, 1997). Indeed, one profile may have higher scores on one dimension and the other profile higher scores on other dimensions and there is no way of judging which of the differences is more important. Although we could concentrate ourselves on only one score of SF36 where a ranking of profiles is feasible, SF36 can not be assumed cardinal.

A preference-based measure derived from SF36

The Sheffield Health Economics Groups (Brazier *et al.*, 1998; Brazier *et al.*, 2002, Brazier *et al.*, 2004) has recently empirically bridged SF36 and utility in order to provide an alternative to existing preference-based measures of health for use in economic evaluation studies such as the EQ-5D (The EuroQol Group, 1990) and the HUI (Feeny *et al.*, 1996). Their approach is

> "to define and value a series of health states using combination of responses levels over SF-36 dimensions. It draws directly from the conceptual and empirical logic of multi-attribute utility theory used in the construction of HUI and EQ-5D where an additive or multiplicative utility function is estimated based on a fractional factorial design from the universe of all possible health states. The "bridge" back to SF36 is formed via the beta coefficients on the utility scoring formula and the corresponding levels on SF36 dimensions."

The derivation of SF6D relies on an algorithm based on six of the eight dimensions of SF36[13]. It has been done for 249 health states valued by 836 respondents from a UK sample. O'Brien *et al.* (2003) have analysed differences between SF6D and the established and widely used utility measure that is HUI. They conclude that it is difficult to disentangle whether differences are due to differences in underlying concepts of health being measured or different utility-theoretic measurement approach. However, SF6D is a valuable addition that permits transforming SF36 into a utility-based measure.

Methodological strategy

On one hand, SF6D has not been applied in any other population except British population, and on the other hand, SF36 is a standardised questionnaire at European

[13]The general health dimension is omitted and role-physical and role-emotional are combined in a unique dimension.

level (Noble *et al.*, 1998). This is the reason for assuming that we can apply the SF6D utility algorithm to the French SF36 available in the 2003 National Health Survey. We use the algorithm[14] based on a consistent version of the model 10 in the paper Brazier *et al.* (2002). This French version of SF6D will represent a reliable cardinal health distribution that can be used to describe the latent variable that determines self-assessed health. The empirical distribution of SF6D in the French population is thus used to scale the intervals of the five (respectively the eleven) categories of self-assessed health in the 2004 IRDES-HHIS.

For every individual, we assume a direct mapping from SF6D to the latent variable subjacent to self-assessed health. An individual's rank according to SF6D, for instance the p^{th} quantile, corresponds to his rank according to self-assessed health in the 2004 IRDES-HHIS. The thresholds, so called c_a, are estimated using a non parametric approach. First, we compute the cumulative frequency of the observations for each of the five (respectively the eleven) categories of self-assessed health. Then, we find the values of the distribution of SF6D that match these frequencies.

The figure 4.7 presents boundaries from SF6D that match the cumulated frequency of the 5-categories self-assessed health and those that match the 11-categories self-assessed health in 2004. In concrete terms, the thresholds of the category a of the French self-assessed health equal the inverse of SF6D empirical distribution F of the cumulative proportion of observations for the category a, i.e. the cumulative value of the upper-bound of the category a. This can be written as:

$$c_a = F^{-1}(G_a) \tag{4.1}$$

Therefore, there are six (respectively twelve) thresholds to consider from $0,337$ (the worse possible status in SF6D) to $0,948$ (the best possible status in SF6D). For the 5-categories self-assessed health, these threshold are $0,337$; $0,364$; $0,457$; $0,574$; $0,717$ and $0,948$ whereas for the 11-categories self-assessed health, they are $0,337$ $0,365$; $0,433$; $0,44$; $0,457$; $0,516$; $0,558$; $0,592$; $0,671$; $0,727$ and $0,948$. The figure 4.7 shows, for instance, that individuals who have reported a health status equal or lower than "good" represent $74,2\%$ of the sample and have a health status lower than $0,717$ according to SF6D. Respectively, the $55,37\%$ of the individuals who have reported a health status equal or lower than 9 have a health status equal or lower than $0,671$ according to SF6D.

[14] Computer programs can be obtained on `www.shef.ac.uk/scharr/sections/heds/mvh/sf-6d/index.html`

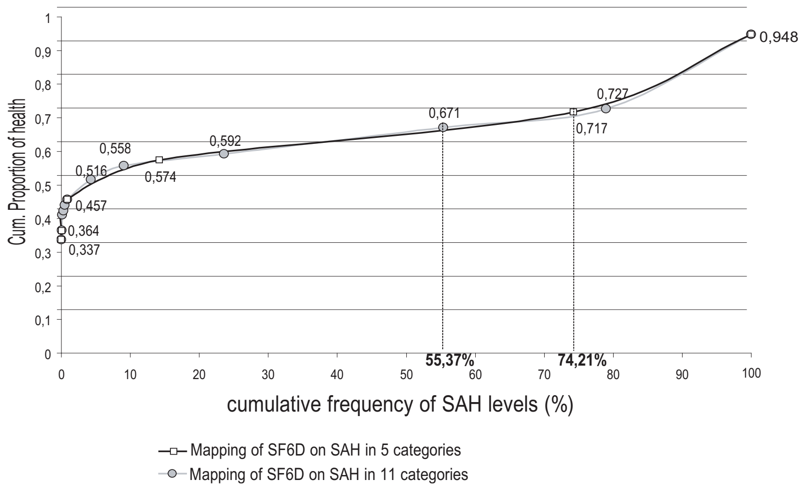

Figure 4.7: Derivation of boundaries from SF6D for self-assessed health described in 5 and in 11 categories (*2004 IRDES-HHIS*)

An interval regression model can then be carried out using these thresholds of the self-assessed health categories and including different regressors to the model. In this context, this is the level of SF6D that is predicted considering that an individual has characteristics Z_i. These characteristics give information on the individual.

We believe that this second indicator can be considered as a benchmark throughout the analysis. Two reasons motivate this assumption. Firstly, interval regression approach outperforms other approaches (van Doorslaer & Jones, 2003; Lecluyse & Clempuut, 2006) on the measurement of inequalities in health and this second indicator is generated on a health distribution in the French general population.

4.3.3 Innovative health index: a first empirical utilisation

The health index generated in chapter 2 is a relevant tool to measure individual health status and could also be a promising tool for the measurement of inequalities in health. We consider it to be another measurement of health for our analysis of inequalities in health in 2004.

The construction of the health index in chapter 2 relies on the 2002 IRDES-HHIS. Nevertheless this construction method can be easily adapted to any other year of the

IRDES surveys. As a result, a new version is thus generated for the 2004 sample data[15]. For the sake of comparability with previous health indicators, the health index is described in the interval[16] $[0; 1]$.

In addition to self-assessed health reported in 5 and 11 categories, each respondent is also assigned a value of the health index. In order to be able to understand the best way to involve this new health index in the analysis of inequalities in health, we have to describe this health index more precisely.

Distributional analysis of the health index

We rely on the empirical case study carried out by Jones *et al.* (2007b, p. 29-49) to give a comprehensive understanding of the health index. The cumulative distribution function for the health index is drawn in figure 4.8 for the full sample. The inverted L-shape of the empirical distribution function emphasises that there is a long left-hand tail which represents relatively few individuals in very bad health. Many people are concentrated in the right-hand tail and so have a higher health index. The vertical line at the right-end of the distribution shows that a large proportion of individuals have a health status which equals 1.

---

[15] Considering the advantages of the interval regression approach, we have tested the possibility to generate the severity weights involved in the construction of the health index within an interval regression using a scaling on SF6D instead of using an ordered Logit.
The difference with the original indicator relies on the introduction of the number of diseases D_{ij}^k of severity k that individual i in household j declares in the interval regression in addition to the vector of individual characteristics Z_i.
In this context, this is the level of the self-assessed health rescaled on the distribution of SF36 that is predicted considering that an individual has characteristics Z_i and D_{ij}^k. The unconditional prediction of the individual regressors vectors $Z_{ij}\beta$ and $D_{ij}\alpha_k$ gives a prediction of each individual's level of SF6D derives from his observed self-assessed health in the IRDES-HHIS survey. Then, the parameter estimates $\widehat{\alpha_k}$ are used as weights to obtain robust self-assessed health indicators in the same way as they were used in the original construction of the health index.
A continuous health index is again generated at individual level by multiplying the number of diseases per severity degree with the associated estimated parameter. Applying the same method as in chapter 2, a new continuous health index is generated.
In this alternative construction, "true" health is based not only on both medical and subjective health dimensions, but also approached by a generic health distribution in the population, namely SF6D. Nevertheless, this SF6D modified health index gives very similar results as the original health index desspite of the lower estimated coefficients for the severity level provided by the interval regression. We have thus decided that this refinement in the methodological construction does not represent an added value to our analysis and have chosen to concentrate on the original construction.

[16] This index on $[0; 1]$ is generated as $I_{ij} = \frac{I_{\max}^{raw} - I_{ij}^{raw}}{I_{\max}^{raw}}$.

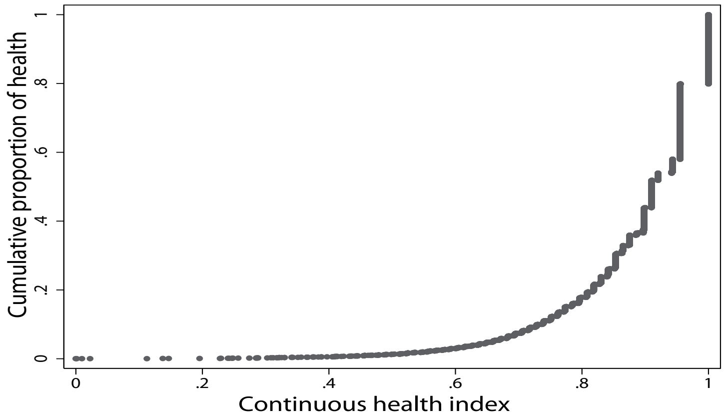

Figure 4.8: Empirical distribution function of the health index.

The quantitative form and the continuous aspect of the health index permit carrying out an ordinary least square regression model. We shall verify if it is appropriate to use a simple linear regression specification with our indicator. We run a simple linear regression on individual characteristics Z_i and analyse residuals from this regression. Figure 4.10 graphically represents the shape of the distribution of residuals.

Obs	8,235
Variance	0.0107124
Skewness	-1.719254
Kurtosis	8.460832

Figure 4.9: Descriptive statistics of residuals of an OLS regression of the health index.

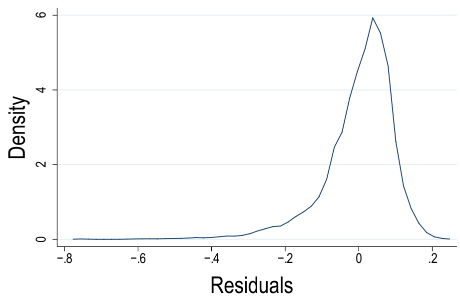

Figure 4.10: Kernel density estimates for OLS residuals

The associated skewness and kurtosis statistics are summarised in table 4.9. These elements show non-normality in the distribution of residuals, which shed some doubt on the use of an OLS regression. This non-normality can be explained by the distribution of the health index, which is truncated at an upper limit of 1. A good way to check if the regression specification is appropriate is to use a reset test. The reset test related to this regression specification is $F(1, 8207) = 12,91$ with a probability of rejection of $Prob > F = 0,0003$. This means that the model is mis-specified and an OLS regression is inappropriate.

Several studies (van Doorslaer & Jones, 2003; Fonseca & Jones, 2003; Lecluyse & Cleemput, 2006) have recently concluded that the interval regression approach outperforms other approaches such as the lognormal distribution or the ordered Probit regression in the measurement of inequalities in health. The health index could thus be used to scale the intervals of self-assessed health. As self-assessed health is involved in the construction of the health index, we can assume that there is a stable mapping from the health index to the latent variable that determines self-assessed health and that this applies for all individuals in the sample. We apply a mapping similar to the one described before with SF6D. The p^{th} quantile of the distribution of the health index corresponds to the p^{th} quantile of the distribution of self-assessed health in the 2004 IRDES-HHIS. The thresholds are estimated using a non parametric approach. The figure 4.11 presents boundaries from the health index that match the cumulated frequency of the 5-categories (respectively the 11-categories) self-assessed health.

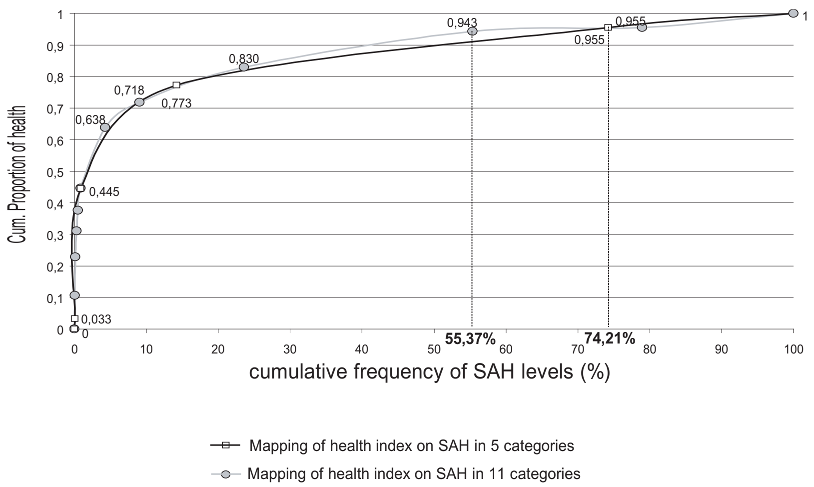

Figure 4.11: Derivation of boundaries from health index for self-assessed health described in 5 and in 11 categories (*2004 IRDES-HHIS*)

The six thresholds to consider for the mapping are 0, 0,033, 0,445, 0,773, 0,955 and 1 (the best possible status for the health index). As for the 11-categories self-assessed health, threshold are 0; 0,107; 0,228; 0,311; 0,376, 0,376, 0,446 0,638; 0,718; 0,83; 0,943; 0,955 and 1. These thresholds are then used in an interval regression model explaining self-assessed health and including various regressors. The health index can thus be used to analyse inequalities in health within an interval regression. We will compare the measurement of inequalities in health offered by this mapping with the measurement obtained by other mappings. This comparative study will allow us to conclude on the empirical use of the health index.

4.3.4 Comparisons of the alternative mappings

When comparing the thresholds from the three different mappings of self-assessed health in five categories, it should be noted that assuming identical health distributions between France and Canada leads to a higher distribution of health status. As illustrated in figure 4.12, there is a lower probability of poor health status with this particular mapping than with the other two.

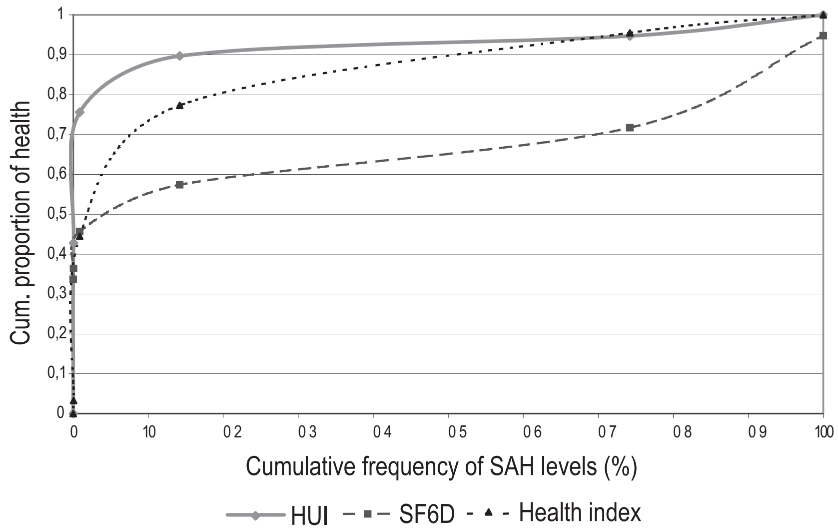

Figure 4.12: Thresholds for self-assessed health in 5 categories according to the mapping.

The mapping with health index describes a distribution of health closer to the mapping with HUI than the one with SF6D. Indeed, applying SF6D thresholds leads to a more compressed health distribution. We supplement this graphical analysis of the mapping by using unilateral tests of *Kolmogorov-Smirnov* on the expected health values of the three health indicators[17]. The results in table 4.6 confirm that the predicted distribution of HUI significantly dominates the predicted distributions of SF6D and the predicted health index. Moreover, the predicted distribution of SF6D is dominated by the predicted distribution of the health index.

[17] Figure 4.12 is based on the predicted health distributions coming from interval regression models using estimated thresholds of the upper and lower bound for each mapping. We carry out a stochastic dominance analysis firstly, based on the lower bound and secondly, based on the upper bound. The first step emphasises that the distribution of SF6D is dominated at first order by the distribution of HUI, but we cannot conclude on stochastic dominance at first order between HUI and the health index as well as between the health index and SF6D. Graphically, it is clear that these curves cross so second order dominance might hold. The analysis on the upper bound shows again that the distribution of HUI significantly dominates the distribution of SF6D. Furthermore, the distribution of the health index also dominates at first order the distribution of SF6D. Still, we cannot conclude on dominance between the distribution of HUI and the health index. To our point of view, this analysis of lower and upper bounds is interesting, nevertheless, as the whole process of the mapping is to make continuous a discrete variable it is worth to supplement it by a dominance analysis based on the predicted distribution of these three health measurements as presented above.

	Predicted Health Index	Predicted SF6D	Predicted HUI
Predicted Health Index		<0,0001***	0,999
Predicted SF6D	1		1
Predicted HUI	<0,0001***	<0,0001***	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.6: P-value of *Kolmogorov-Smirnov* test related to the three predicted health measurements on self-assessed in 5 categories

The same differences are observed when self-assessed health is coded in eleven categories: the mapping with the health index gives significantly dominates at first order the distribution of SF6D as shown in figure 4.13 and in table 4.7.

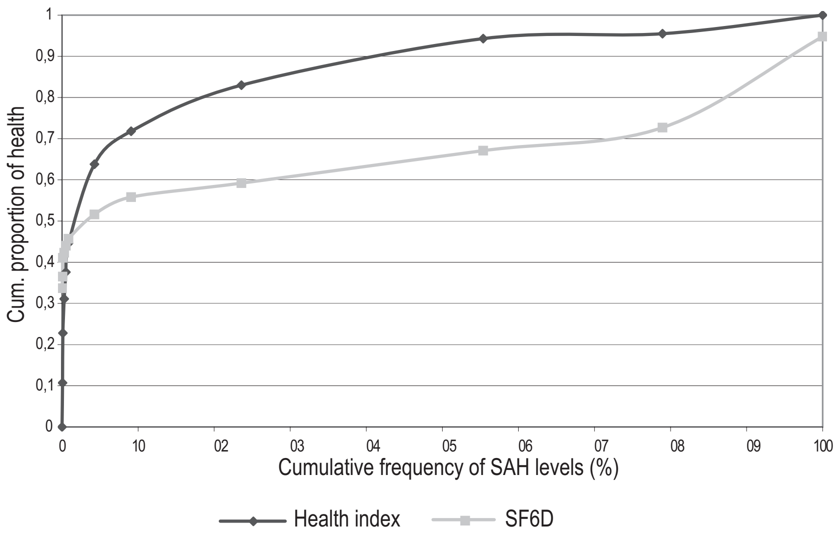

Figure 4.13: Thresholds for self-assessed health in 11 categories according to the mapping.

	Predicted Health Index	Predicted SF6D
Predicted Health Index		<0,0001***
Predicted SF6D	1	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 4.7: P-value of *Kolmogorov-Smirnov* test related to the predicted health measurements on self-assessed in 11 categories

All these thresholds represent the assumed distribution of health in France in 2004. As they exhibit different distributions of health, it is expected to observe differences in the measurement of inequalities in health from one distribution of health to another. Intuitively, lower thresholds might lead to a higher inequality; the following comprehensive analysis of inequalities in health will allow us to go further than this intuition in the comparisons of these mappings.

4.4 Measuring income-related inequality in health

4.4.1 Measurement method

Our study of inequalities in health relies on the calculation of concentration indices as presented in chapter 3. Concentration indices capture the socioeconomic dimension of health inequalities and use information from the whole distribution of health over income (Jones *et al.*, 2007a). The analysis controls for various covariates of health such as demographic, socioeconomic and health insurance characteristics. Nevertheless, it is important to underline that the study does not allow any causal interpretation; regression coefficients in particular may reflect either reverse causality or joint determination due to unobserved factors.

We assume that a linear regression model defines the health status y_i of individual i according to k regressors, such as $k = (1, ..., K)$. This can be written:

$$y_i = a + \sum_{k=1}^{K} b_k x_{ki} + \epsilon_i \tag{4.2}$$

The random error term, ϵ_i is assumed to have expected mean value equal to zero and constant variance. The b_k are assumed constant for every individual i.

The concentration index requires a ranking variable for the population. We use the logarithm of the equivalent household income per consumption unit as described in subsection 4.2.2. The concentration index related to this measure of health on income is given by the following equation

$$CI = (\frac{2}{\overline{y}})cov(y_i, R_i) \tag{4.3}$$

where R_i is the cumulative proportion of the population ranked by increased income up to individual i and $\overline{x} = E(x_i)$.

The linear regression model includes several regressors, namely age-gender categories, levels of education, categories of activity status, socioeconomic status and health insurance variables. These latter variables indicate whether the individual is covered by private health insurance beyond compulsory insurance or by the CMU. Marital status was firstly also involved in preliminary analyses but it has been dropped for non significance.

Sample characteristics

The table 4.8 presents some descriptive statistics of the 2004 IRDES-HHIS sample. The mean value of equivalent income is 1,986 euros per month. This value is tricky to compare to overall French statistics, as this value concerns a specific sample and is in gross salary terms. For instance, the mean value of net income for the 18-59 years old was about 17,879 euros per year in 2004[18]. Nevertheless, it has been shown that IRDES-HHIS surveys under-estimate the average income as regard to national accounts (Grignon, 2003). Regarding unemployment status, it represents almost 6% of the sample. In reality, the unemployment rate equals 9.1% in December 2004[19]. Our sample, once again, presents again lower proportions than proportions observed in national statistics. These differences are explained by both our restricted age sample and the inability of IRDES to interview precarious households. As for the supplementary health insurance, the IRDES-HHIS under-estimates the proportion of CMU beneficiaries because the proportion in the metropolitan France equals 7% in 2005 (Boisguérin, 2005). This under-estimation is due to an under-representation of precarious people in most of the general population surveys.

Variables	Mean
Income per CU (€/month)	1985,90
Age	38,46
Education less	45,45
Education 2	21,14
Education 3	33,41
Private health insurance	90,69
No private health insurance	7,37
CMU	1,94
Employed	67,83
Inactive	1,15
Homemaker	5,34
Retired	5,84
Unemployed	5,85
Student	13,98
Employee	26,98
Farmer	1,99
Self-employed	4,80
Executive	14,61
Technician	22,27
Skilled worker	21,07
Unskilled worker	7,67

Table 4.8: Descriptive statistics (*2004 IRDES-HHIS*).

4.4.2 Explaining health within a linear model

As discussed in section 4.4.1, we specify and estimate a linear regression model explaining self-assessed health. We carry out five different interval regression models using the five alternative mappings. It is useful to stress that these regression models do not provide a structural model for health and therefore the estimates do not give a causal interpretation. However, these models might be interpreted as reduced form static models of demand for health whose estimates provide an indication of how exogenous changes

[18]INSEE `www.insee.fr`, La France en faits et en chiffres.
[19]INSEE `www.insee.fr`, La France en faits et en chiffres.

in health determinants can affect the degree of income-related inequality in health (Garcia & Lopez, 2007). Moreover, coefficients of interval regression models are measured on the same scale as the cut-points so they can be interpreted in terms of changes in the distribution of health (Jones *et al.*, 2007a).

The table 4.9 presents results of the interval regressions of the five different mappings. It is noteworthy that relationships are qualitatively and significatively similar regardless of the mapping or the scale of self-assessed health. There is, nevertheless, a number of cases where they differ. We will consider these after describing their similarities.

Coefficients of the HUI mapping tend to be lower than those concerning the two other mappings. These lower values confirm the stochastic dominance at first order of HUI on the two other mappings. It could also be explained just as it is proposed in Leu and Schellhorn (2006), who compare three different scalings for self-assessed health and observe that coefficients depend on the spreading of the health distribution involved in the mapping. There is indeed a direct consequence of the less compressed health distribution of the HUI as opposed to the two others distributions.

As expected, health decreases substantially with age for both genders. For instance, women as well as men between 56 and 65 years old on average report a health status lower than men aged 36 to 45 years old. With HUI, their health status is around 0,02 times lower, whereas with the SF6D mapping (respectively the health index mapping) regardless of the self-assessed health scaling, their health status is around 0,06 (respectively 0,07) times lower for women and 0,05 times lower for men. Health is likely to be worse for women than men. Incidentally, there is no significant effect on health of women aged 26 to 35 years old, this lack of significance could be explained by a better assessment of health status of this age category.

When it comes to socioeconomic characteristics, income has a positive and significant direct effect on health regardless of the health indicator. Similarly, more educated people, i.e those having a primary/secondary or high school level education, have a significantly better health than less educated people irrespective of the health indicator. Compared to the employed people, homemakers, inactive and unemployed people have a negative and significant lower health, irrespective of the health mapping of a self-assessed health in 5 categories.

Variables	Predicted HUI			Predicted SF6D on 5-SAH			Predicted health index on 5-SAH			Predicted SF6D on 11-SAH			Predicted health index on 11-SAH		
	Coeff.	S.E	P-value	Coeff.	S.E	P-value	Coeff.	S.E	P-value	Coeff.	S.E	P-value	Coeff.	S.E	P-value
F 16-25	0,0085	0,0021	0	0,0229	0,0051	0	0,0247	0,0057	0	0,0290	0,0047	0	0,0344	0,0060	0
F 26-35	-0,0004	0,0018	0,847	0,0013	0,0043	0,759	0,0005	0,0049	0,923	0,0060	0,0040	0,130	0,0085	0,0052	0,100
F 36-45	-0,0077	0,0018	0	-0,0165	0,0042	0	-0,0201	0,0048	0	-0,0093	0,0039	0,016	-0,0119	0,0050	0,018
F 46-55	-0,0165	0,0018	0	-0,0405	0,0042	0	-0,0482	0,0049	0	-0,0345	0,0039	0	-0,0485	0,0051	0
F 56-65	-0,0239	0,0023	0	-0,0552	0,0053	0	-0,0682	0,0063	0	-0,0474	0,0049	0	-0,0695	0,0065	0
M 16-25	0,0157	0,0021	0	0,0434	0,0049	0	0,0424	0,0055	0	0,0489	0,0045	0	0,0516	0,0058	0
M 26-35	0,0066	0,0018	0	0,0191	0,0042	0	0,0186	0,0048	0	0,0149	0,0039	0	0,0160	0,0051	0,002
M 36-45	ref.														
M 46-55	-0,0123	0,0018	0	-0,0311	0,0041	0	-0,0352	0,0048	0	-0,0247	0,0038	0	-0,0291	0,0050	0
M 56-65	-0,0195	0,0023	0	-0,0476	0,0054	0	-0,0558	0,0063	0	-0,0414	0,0049	0	-0,0544	0,0065	0
Log income	0,0050	0,0008	0	0,0110	0,0019	0	0,0136	0,0022	0	0,0069	0,0017	0	0,0119	0,0023	0
Education less	ref.														
Education 2	0,0040	0,0012	0,001	0,0083	0,0028	0,002	0,0111	0,0032	0	0,0052	0,0026	0,042	0,0110	0,0033	0,001
Education 3	0,0055	0,0012	0	0,0135	0,0028	0	0,0159	0,0032	0	0,0068	0,0026	0,009	0,0137	0,0033	0
Private health insurance	0,0038	0,0016	0,021	0,00423	0,0038	0,026	0,0091	0,0043	0,034	0,0110	0,0035	0,002	0,0186	0,0045	0
No private health insurance	ref.														
CMU	-0,0113	0,0035	0,001	-0,0209	0,0080	0,009	-0,0289	0,0094	0,002	-0,0068	0,0075	0,359	-0,0118	0,0098	0,228
Employed	ref.														
Inactive	-0,0264	0,0040	0	-0,0403	0,0090	0	-0,0688	0,0105	0	-0,0360	0,0085	0	-0,0752	0,0111	0
Homemaker	-0,0056	0,0020	0,004	-0,0111	0,0045	0,014	-0,0165	0,0053	0,002	-0,0040	0,0042	0,339	-0,0070	0,0055	0,202
Retired	0,0007	0,0023	0,772	-0,0003	0,0052	0,949	0,0003	0,0062	0,96	-0,0034	0,0048	0,475	0,0042	0,0064	0,514
Unemployed	-0,0044	0,0019	0,020	-0,0103	0,0043	0,016	-0,0141	0,0050	0,005	-0,0083	0,0040	0,038	-0,0097	0,0052	0,063
Student	0,0006	0,0018	0,736	0,0041	0,0043	0,347	-0,0006	0,0048	0,894	0,0032	0,0040	0,434	-0,0053	0,0051	0,300
Employee	ref.														
Farmer	0,0043	0,0031	0,166	0,0075	0,0072	0,299	0,0103	0,0083	0,217	0,0109	0,0067	0,103	0,0182	0,0087	0,036
Self-employed	0,0042	0,0021	0,047	0,0137	0,0049	0,006	0,0114	0,0056	0,043	0,0136	0,0046	0,003	0,0138	0,0059	0,020
Executive	0,0033	0,0016	0,04	0,0096	0,0037	0,011	0,0091	0,0043	0,032	0,0078	0,0034	0,024	0,0097	0,0045	0,030
Technician	0,0030	0,0013	0,021	0,0065	0,0030	0,032	0,0081	0,0035	0,02	0,0082	0,0028	0,003	0,0117	0,0036	0,001
Skilled worker	-0,0018	0,0013	0,184	-0,0041	0,0031	0,191	-0,0049	0,0036	0,174	0,0002	0,0029	0,937	-0,0012	0,0038	0,744
Unskil. worker	-0,0029	0,0018	0,095	-0,0092	0,0040	0,023	-0,0089	0,0047	0,058	-0,0029	0,0038	0,442	-0,0051	0,0049	0,294
Constant	0,8864	0,0061	0	0,5770	0,0142	0	0,7644	0,0163	0	0,5938	0,0131	0	0,7727	0,0171	0

Table 4.9: Health interval regressions coefficients (*2004 IRDES-HHIS*).

As regard to the very weak proportion of inactive people in the sample (1,15%), a poor health might be the reason for inactivity. It could also be an illustration of the justification bias of inactivity as described in Kerkhofs and Lindeboom (1995). The authors show that people are likely to report a poor health status to justify their inactivity. Nevertheless, unemployment and inactivity are associated with an excess mortality for both men and women among individuals aged 16-65 years old. Indeed, this result is in line with recent findings which show that during the five years following an unemployment period, the annual risk of death for an unemployed individual is *ceteris paribus* approximately three times higher than that of the general 16-60 population (Mesrine, 2000). Moreover, health status of people who are unemployed is significantly worse than that of people who are employed because unemployed people have significantly higher rates of psychosocial diseases such as anxiety and depression (Khlat & Sermet, 2004). There is no significant effect on health of being retired or being a student. Unskilled workers report a poorer health status compared with the reference group of employees when self-assessed is considered on a five categories scale. Self-employed people, executive people and technicians are in significantly better health than employees, whatever the mapping of health.

The effect of private health insurance appears to be positively related to health. Irrespective of the health mapping corresponding to a 5-categories self-assessed health, a negative relationship links health and CMU. It is due to the CMU eligibility conditions which target individuals with very low incomes and often imply low health statuses, too. Similarly, Boisguérin (2005) shows that individuals tend to enrol CMU if they anticipate health care needs; there is thus a selection bias.

Mismatches in the results become particularly obvious when we consider self-assessed health described in eleven categories. These differences concern a loss of significance of some regressors, such as being a beneficiary of CMU, being homemakers or being an unskilled worker. When coded on eleven categories, self-assessed health seems to be less correlated to particular socioeconomic variables but a significant correlation is still observed with the log of income. This can be explained by these eleven categories which smooth away of correlations between health and regressors due to a larger range of possible reported health statuses. There can also be differences in a significance gain with farmers, who are in significantly better health than employees according to both mapping on the 11-categories self-assessed health. It is also remarkable that when all the mappings are non significant, there can be some disparities in the signs of the relationship with health. This is the case for students or skilled workers. Nevertheless, as the degree of significance is highly exceeded, we believe that these differences are of little importance.

4.4.3 Global concentration indices: income-related inequality in health

Prior to the decomposition of inequalities in health, we can analyse the global concentration index of inequality in health in 2004 according to the mapping and to the scale of self-assessed health. This global health concentration index measures the income-related inequalities in health, which is the prime goal of our analysis.

The table 4.10 recapitulates the value of the total health inequality.

Health indicators	CI	Newey-West S.E	[95% Confidence Interval]
Predicted HUI	0,00194	0,00010	[0,00175; 0,00214]
Predicted SF6D on 5-SAH	0,00541	0,00034	[0,00474; 0,00609]
Predicted health index on 5-SAH	0,00566	0,00029	[0,00508; 0,00624]
Predicted SF6D on 11-SAH	0,00195	0,00031	[0,00134; 0,00253]
Predicted health index on 11-SAH	0,00431	0,00029	[0,00374; 0,00489]

Table 4.10: Concentration indices for income-related health inequality in 2004 (*2004 IRDES-HHIS*).

The five concentration indices related to the predicted health indicator are positive and describe an inequality in health favouring the richest individuals. Quantitatively, some differences are shown. When the 5-points self-assessed health is cardinalised with the HUI, the value of the health concentration is lower than any other mappings with the same self-assessed health variable. This difference in magnitude can be explained by the higher thresholds of HUI. The HUI distribution of health gives a lower probability to poor health statuses. Conversely, the health index describes well poor health statuses and this feature is illustrated by the higher value of the health concentration index related to this mapping, regardless of the scale on which self-assessed health is described. Similarly, the thresholds associated to SF6D are low, which explains the higher health concentration index. It is remarkable that with the 11-points self-assessed health, the inequality in health decreases. This wider scale implies a moving of individual health reports over the larger scale and therefore a lower concentration in extreme categories. Nevertheless, it is worth to stress that the mapping using SF6D is the trickiest one. Indeed, the income-related inequality associated with this mapping substantially changes with the scale of self-assessed health. When self-assessed health is described on a 5-points scale, the value of the health concentration related to SF6D is similar to the value obtained with a mapping using the health index. However, when self-assessed health is described on 11-categories, the SF6D mapping leads to a value of concentration index closer to the concentration index with HUI. The inequality in health seems to be sensitive both to the mapping and to the number of responses categories of self-assessed health. These differences between mappings and scale may be clearer with a decomposition of the inequality in health. It is thus of interest to go further than this synthetical measurement. Moreover, this decomposition will allow us to disentangle the part of inequitable inequalities in this global measurement of income-related inequality in health.

4.5 Explaining income-related inequality in health

4.5.1 Measurement method

An attractive feature of the concentration index is its ability to be decomposed into contributions of each of the regressors involved in the econometric model for health (Wagstaff *et al.*, 2003). If we substitute the concentration index formula described in equation 4.3 in the expression of the regression linear model (equation 4.2), we obtain

$$CI = \sum \left(b_k \frac{\overline{x_k}}{\overline{y}} \right) CI_k + \frac{2}{\overline{y}} cov(\epsilon_i, r_i) \tag{4.4}$$

The concentration index is thus assumed to be made up of two components: an explained component equal to a weighted sum of the concentration indices of the k regressors, and a residual one. The residual component reflects the health inequality which is not explained by systematic variation across income groups in the regressors. In the case of the interval regression approach, no residuals can be computed and the decomposition reduces to the explained part of the previous equation. The use of interval regression is more efficient than standard methods of ordered Probit or Logit. Therefore the linear index $z_i\beta$ gives a measure of predicted utility from an individual i, who has characteristics Z.

In this way, the estimated health inequality can be simply written

$$\widehat{CI} = \sum_k \widehat{\eta_k} CI_k \tag{4.5}$$

Therefore, the decomposition method separates the contribution of each regressor k into two quantifiable elements: its impact on health, as measured by the health elasticity $\widehat{\eta_k}$, and the degree of inequality of its own distribution with respect to income, as measured by the income concentration index CI_k.

4.5.2 Concentration indices over income

The first step of the decomposition method allows us to analyse the concentration indices of each regressor over the income distribution. The second column in table 4.11, called CI, shows the distribution over income of each regressor involved in the regression model explaining health.

Variables			Predicted HUI			Predicted SF6D on 5-SAH			Predicted health index 5-SAH			Predicted SF6D on 11-SAH			Predicted health index 11-SAH		
	Mean	CI	Elast.	Contrib.	%	Elast.	Contrib.	%	Elast.	Contrib.	%	Elast.	Contrib.	%	Elast.	Contrib.	%
F 16-25	0,108	-0,173	0,001	-0,0002	-8,82%	0,004	-0,0006	-11,94%	0,003	-0,0005	-9,36%	0,005	-0,0008	-42,14%	0,004	-0,0007	-16,92%
F 26-35	0,110	-0,020	0,000	0,0000	0,04%	0,000	0,0000	-0,08%	0,000	0,0000	-0,02%	0,001	0,0000	-1,02%	0,001	0,0000	-0,49%
F 36-45	0,116	-0,078	-0,001	0,0001	3,81%	-0,003	0,0002	4,12%	-0,003	0,0002	3,67%	-0,002	0,0001	6,47%	-0,002	0,0001	2,82%
F 46-55	0,111	0,094	-0,002	-0,0002	-9,57%	-0,007	-0,0006	-11,77%	-0,006	-0,0006	-10,22%	-0,006	-0,0005	-28,03%	-0,006	-0,0006	-13,32%
F 56-65	0,068	0,188	-0,002	-0,0003	-16,98%	-0,006	-0,0011	-19,68%	-0,005	-0,0010	-17,69%	-0,005	-0,0009	-47,23%	-0,005	-0,0010	-23,39%
M 16-25	0,113	-0,141	0,002	-0,0003	-13,93%	0,007	-0,0010	-19,32%	0,005	-0,0008	-13,74%	0,008	-0,0012	-60,80%	0,007	-0,0009	-21,68%
M 26-35	0,096	0,020	0,001	0,0000	0,72%	0,003	0,0001	1,04%	0,002	0,0000	0,73%	0,002	0,0000	2,25%	0,002	0,0000	0,82%
M 36-45	ref.																
M 46-55	0,099	0,120	-0,001	-0,0002	-8,10%	-0,005	-0,0006	-10,29%	-0,004	-0,0005	-8,48%	-0,004	-0,0004	-22,84%	-0,003	-0,0004	-9,11%
M 56-65	0,062	0,186	-0,001	-0,0002	-12,43%	-0,004	-0,0008	-15,20%	-0,004	-0,0007	-12,98%	-0,004	-0,0007	-36,97%	-0,004	-0,0007	-16,41%
Log income	7,434	0,041	0,040	0,0016	84,45%	0,123	0,0051	93,64%	0,116	0,0048	84,61%	0,078	0,0032	163,94%	0,100	0,0041	95,44%
Education less	ref.																
Education 2	0,211	-0,032	0,001	0,0000	-1,51%	0,003	-0,0001	-1,58%	0,003	-0,0001	-1,53%	0,002	-0,0001	-2,74%	0,003	-0,0001	-1,96%
Education 3	0,334	0,242	0,002	0,0005	24,58%	0,007	0,0016	30,44%	0,006	0,0015	26,15%	0,003	0,0008	42,71%	0,005	0,0013	29,10%
Private health insurance	0,907	0,041	0,004	0,0001	7,70%	0,006	0,0002	4,34%	0,010	0,0004	6,83%	0,015	0,0006	31,41%	0,019	0,0008	18,06%
No private health insurance	ref.																
CMU	0,018	-0,711	0,000	0,0002	7,94%	-0,001	0,0004	7,41%	-0,001	0,0004	7,44%	0,000	0,0001	6,75%	0,000	0,0002	3,93%
Employed	ref.																
Inactive	0,012	-0,303	0,000	0,0001	5,12%	-0,001	0,0002	3,92%	-0,001	0,0003	4,88%	-0,001	0,0002	9,80%	-0,001	0,0003	6,91%
Homemaker	0,053	-0,327	0,000	0,0001	5,49%	-0,001	0,0003	5,39%	-0,001	0,0003	5,86%	0,000	0,0001	5,47%	0,000	0,0001	3,24%
Retired	0,059	0,184	0,000	0,0000	0,39%	0,000	0,0000	-0,10%	0,000	0,0000	0,07%	0,000	-0,0001	-2,87%	0,000	0,0001	1,18%
Unemployed	0,058	-0,294	0,000	0,0001	4,15%	-0,001	0,0003	4,92%	-0,001	0,0003	4,90%	-0,001	0,0002	11,05%	-0,001	0,0002	4,38%
Student	0,140	-0,170	0,000	0,0000	-0,81%	0,001	-0,0001	-2,71%	0,000	0,0000	0,31%	0,001	-0,0001	-5,86%	-0,001	0,0001	3,31%
Employee	ref.																
Farmer	0,020	-0,285	0,000	0,0000	-1,36%	0,000	-0,0001	-1,18%	0,000	-0,0001	-1,19%	0,000	-0,0001	-4,81%	0,000	-0,0001	-2,72%
Self-employed	0,048	-0,034	0,000	0,0000	-0,39%	0,001	0,0000	-0,63%	0,001	0,0000	-0,38%	0,001	0,0000	-1,75%	0,001	0,0000	-0,60%
Executive	0,146	0,471	0,001	0,0002	12,54%	0,002	0,0010	18,34%	0,002	0,0007	12,74%	0,002	0,0008	41,85%	0,002	0,0008	17,58%
Technician	0,223	0,211	0,001	0,0002	7,76%	0,002	0,0005	8,45%	0,002	0,0004	7,69%	0,003	0,0006	29,86%	0,003	0,0006	14,50%
Skilled worker	0,211	-0,222	0,000	0,0001	4,63%	-0,001	0,0003	5,30%	-0,001	0,0003	4,63%	0,000	0,0000	-0,84%	0,000	0,0001	1,51%
Unskil. worker	0,077	-0,366	0,000	0,0001	4,57%	-0,001	0,0004	7,17%	-0,001	0,0003	5,08%	0,000	0,0001	6,33%	0,000	0,0002	3,80%
Total CI				0,0019			0,0054			0,0057			0,0019			0,0043	
CI*				-0,0013			-0,0045			-0,0039			-0,0045			-0,0042	
I=CI-CI*				0,0032			0,0099			0,0095			0,0064			0,0085	

Table 4.11: Decomposition of concentration indices for health (*2004 IRDES-HHIS*).

The concentration indices for the determinants of health are identical for all the health indicators as the inequality is measured over the same ranking variable. This ranking variable is the equivalent household income, which is considered in log as this simple transformation presents advantages for residuals and ranks individuals in the same way. As expected, the concentration index obtained for the log of income is lower than the Gini index calculated in section 4.2.2.

With respect to the age-gender categories, it is clear, regardless of gender, that the youngest are concentrated in lower income groups whereas people over 46 years old are concentrated in higher income groups. Unlike their male peers, middle-aged women appear to be poor. Moreover, it is remarkable that there is an inequality in income favouring men: when similar pattern is observed, concentration indices over income are most of the time more favorable for men than women. The most-educated individuals are heavily concentrated in the richest income groups. When people have a primary/secondary school level education, they are also concentrated in the richest income groups but the value of the concentration index associated is very weak as compared to the concentration index for people having at least baccalauréat, i.e A-levels.

Homemakers, students, inactive and unemployed people are concentrated in lower income groups, the most disadvantaged being homemakers. When it comes to retired people, an inequality favouring the richest is observed. As the sample only includes individuals between the age of 16 and 65, we can presume that those who have retired earlier have either done so for a reason of poor health or because it was economically more advantageous. Nevertheless, it is clear that the needy people, even when they have a poor health status, are likely to keep on working.

As regard to social status, except executives and technicians who belong to higher income groups, all the other social statuses are concentrated among lower income groups. In particular, farmers and unskilled workers experienced the highest inequality over income.

Finally, concentration indices concerning health insurance accord with primary intuition. Having a supplementary health insurance is widespread in the population; the concentration index associated is weak but favours higher incomes. Indeed, some analyses on the IRDES-HHIS show that those who have no supplementary health insurance are either the youngest, who are healthy and have a lowest preference for health, or the poorest who cannot pay for it, or else old women who were beneficiary of their husband's cover and do not subscribe after widowhood. As for CMU, it appears to be the highest concentration index. It strongly favours the poorest because it is means-tested.

4.5.3 Contribution to the income-related inequality in health

The decomposition method previously presented gives the contribution of each regressor to the income-related inequality in health in 2004. We now move to the explanation of the inequality in health according to the health mapping and the regressors. Table 4.11 exhibits the contribution to the income-related inequality in health of each regressor for each mapping. This contribution value is presented in exact value and in proportion of the total inequality[20]. Table 4.11 shows that from one mapping of health status to another, regressors mainly contribute in the same way to the inequality. Nevertheless, it is remarkable that the three mappings on the self-assessed health in five categories give similar contributions to inequality in health whereas those on the self-assessed health in eleven categories are very different. Differences in magnitude have already been underlined in the literature. When comparing the scaling of Flemish self-assessed health using the Flemish EQ-index and the scaling with the Canadian HUI in a perspective of measurement of inequalities in health, Lecluyse & Cleemput (2004) show different values in terms of magnitude of concentration indices.

Despite these mismatches, irrespective of the health indicator, the highest contributions come from the same regressors: log of income, higher education, older age-gender categories, higher social status such as executive or technician. In the first subsection, we consider regressors that contribute the most to the total inequality, and in the second section, we focus on regressors whose contributions strongly vary according to the mapping.

Some regressors explain most of the total inequality

The contribution of income to inequality in health is relevant. Regardless of the mapping, its contribution to inequality is at least three times higher than the contribution of any other parameter. There is also a high positive elasticity of health with income. This result is in line with most of the European analyses on income-related inequalities in health. For instance, in Switzerland, the contribution of income to inequality is around 60% (Leu & Schellhorn, 2006) and in Spain, it equals 102.5% or 30.6% according to the mapping.

People with more years of schooling tend to have better health. Education interacts in many ways with income and having a higher education level is the second most explicative parameter of inequality. There are several references in literature which have emphasised the protective role played by a higher education level on mental and physical health (Feinstein *et al.*, 2006) or mortality (Kunst & Mackenbach, 1994).

Some age-classes comprehensively contribute to inequality. It is the case of older people, especially women. Their contribution to inequality is negative, decreasing the income-related inequality in health. This reduction comes from the fact that older people are both richer as shown by the associated positive concentration index over income and in

[20] The ratio of the contribution of each regressor by the health concentration index reminded as total CI in the last rows of table 4.11 gives the percentage of this contribution in the total inequality.

worse health as shown by the negative elasticity of health with older age-gender categories. Asymmetrically, the elasticity of health with women aged 16-25 is positive, and the high negative contribution of younger women to inequality is due to their concentration in low income levels in spite of their good health status.

Being executive or technician explains the inequality in health in a similar proportion than having a high education level. Individuals belonging to these social statuses enjoy a better health status, which is shown by the positive elasticity with health. This result is in line with other analyses of inequalities in health according to socioeconomic status, using other health indicators, such as mortality or specific diseases (Mackenbach, 2006).

Some regressors are particularly sensitive to the mapping or the scale of self-assessed health

It is noteworthy that contributions to inequality are qualitatively similar regardless of the mapping. There are some exceptions with characteristics of activity status and socioeconomic status. For example, being retired and being a skilled worker. Indeed, whereas all the other mappings show a positive contribution to the inequality of these characteristics, the mapping using SF6D with the 11-categories self-assessed health describes a negative contribution. Similarly, being a student always contributes for a reduction of the inequality level, except when self-assessed health is mapped using the health index. Nevertheless, it is noticeable that these unsteady variations concern individual variables, which are weakly contributing to the inequality.

As regard to differences in the magnitude of the contributions, we have previously mentioned that the 11-categories self-assessed health generally presents contributions doubling the contributions of the other scale. Nevertheless, the opposite case can also be observed for this scaling. In this context, it is interesting to underline the contribution of CMU however it is lower than other major explicative regressors. It contributes for about 8% when self-assessed health is mapped on five categories and for only 4% when self-assessed health is mapped with health index on eleven categories. Nevertheless, CMU always contributes positively to the inequality in health. As the concentration index of CMU-beneficiary over income was negative, the positive contribution is due to a negative relationship between health and asking for CMU: the sickest often are also the poorest, which increases their will to ask for CMU.

4.5.4 Legitimate or illegitimate income-related inequalities in health?

So far we have measured a concentration index of the income-related inequality in health, which does not distinguish policy relevant and policy irrelevant variables. A variable is considered as policy irrelevant if it is impossible to alter either its direct effect on health or its joint distribution with income. The effects of such policy irrelevant variables have to be removed from the income-related health inequality in order to evaluate the level of inequity in health. The distinction between the two types of variables relies on the policy context. However, the literature mainly considers age and gender as policy irrelevant variables[21] (Gravelle , 2003) and a standardisation on age and gender is carried out in most of the economic and epidemiological analyses (van Doorslaer & Koolman, 2004; Gravelle & Sutton, 2003; Boissonnat & Mormiche, 2007). Kakwani *et al.* (1997) refer about legitimate inequalities and argue that variations in health due to biological differences can be considered to a large degree legitimate.

As we have seen, the contribution of age and sex categories to the income-related health inequality is far from negligible, particularly for extreme age-classes. We can thus expect differences in the results if we remove the effects of these policy irrelevant variables from the income-related inequality.

There are two methods for standardisation: the direct and indirect method. The direct method determines the distribution of health that would be observed if every individual had the same age and gender structure. The policy irrelevant variables are thus fixed at a reference level which is the same for all individuals. As for the indirect method, it represents the difference between actual health and expected health, where expected health for an individual is the average health of individuals with the same of age and gender characteristics as him.

Gravelle (2003) shows that the indirect standardisation leads to inconsistent estimates of income-related inequality in health and recommends the direct standardisation method. The direct standardisation is also advisable because it relies on full information on the policy relevant and policy irrelevant variables affecting health.

We implement a direct standardisation of the previous concentration indices on age and gender. The three last rows of table 4.11 describe the calculation of inequity in health. Regardless of the mapping, the income-related inequity in health is higher than the income-related inequality in health. Again, the value of inequity in health is lower when self-assessed health is scaled on HUI. It is remarkable that the inequity value when the 11-categories self-assessed health is mapped with SF6D is less similar to the inequity value when self-assessed health is mapped with HUI, conversely to the previous similarity in the total CI.

Our analysis shows the existence of income-related inequalities in health in France in 2004 and underlines that some social individual characteristics, such as income, social

[21] Gravelle (2003) underlines that demographic factors could even be considered as policy relevant factors as it may be possible to alter the joint distribution of age and income by for example a taxation policy or to change the relationship between age and health by targeting health care towards elderly. The distinction between policy relevant and policy irrelevant variables can be linked to the distinction between individual characteristics coming from responsibility and those coming from circumstances. Age and gender are individual characteristics that are independent of individual responsability.

status and education, are explaining a large part of these inequalities. The magnitude of inequalities in health in 2004 arouses the interest to look at the changes over time in the income-related inequalities in health. Specifically, even if the contribution of CMU to the inequality in health is weak, it is of interest to analyse if its introduction has influenced the contributions of other social variables to the inequality. As a result, the following section carries out a comparative analysis of income-related inequalities in health between 1998 and 2004 and use an innovative decomposition method to understand the changes over time.

4.6 Income-related inequality in health in 1998: a comparison with 2004

We shall firstly carry out the analysis of income-related inequalities in health on the 1998 IRDES-HHIS and then compare results with those in 2004. Wagstaff *et al.* (2003) decompose differences in inequality over time, using the well known Oaxaca decomposition whereby differences between the concentration indices at period t and at period $t-1$ can be written as a sum of changes in concentration index, weighted by health elasticities and changes in health elasticities weighted by concentration indices of respective regressors. Considering that this method mainly relies on changes in concentration indices and changes in health elasticities of each regressor, we think that it is insightful to limit ourselves to an analysis of these changes.

4.6.1 Measurements of health in 1998

In 1998, self-assessed health is available in eleven categories only. The measurement of health in 1998 is analogous to the measurement of health in the 2004 data set. Again, we use the empirical distributions of SF6D and the health index are thus used to scale the intervals of the eleven categories of self-assessed health in the 1998 IRDES-HHIS as it has been done with the 2004 data set. The figure 4.14 presents boundaries from SF6D and from health index that match the cumulated frequency of each self-assessed health category in 1998. We then carry out an interval regression model using the thresholds of the self-assessed health categories and including various regressors to the model.

The comparative analysis thus relies on the measurement of inequalities with two different mappings of the 11-categories self-assessed health at two points of the time.

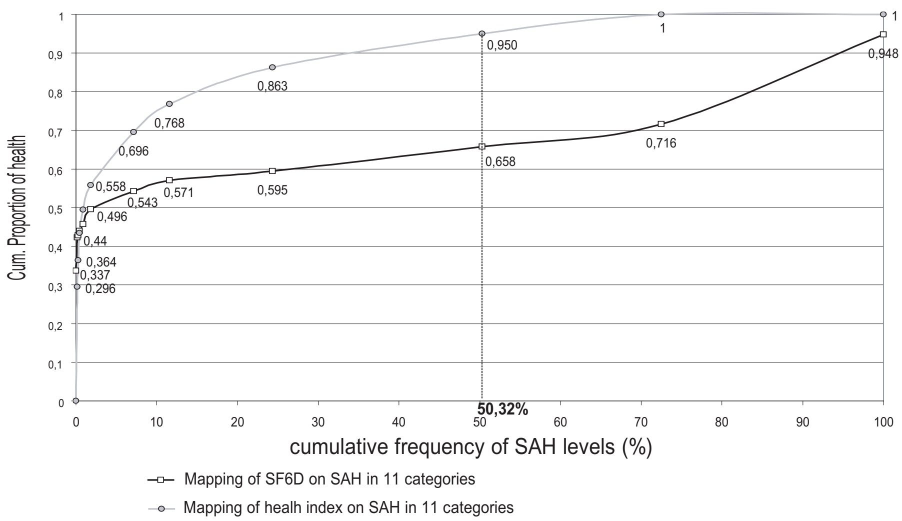

Figure 4.14: Derivation of boundaries from SF6D and the health index for self-assessed health described in 11 categories (*1998 IRDES-HHIS*)

4.6.2 Data and variables

Our comparative analysis is a cross-sectional analysis using the same regressors variables, except CMU. Indeed, the CMU reform has been introduced in 2000. Before this, the poorest individuals had AMG, which was fairly limited, mainly exempting individuals from having to pay the *ticket modérateur*. The table 4.12 presents some descriptive statistics of samples in the two years.

Variables	Mean	
	1998	2000
Income per CU (€/month)	1499, 01	1985,90
Age	37,73	38,46
Education less	48,62	45,45
Education 2	22,14	21,14
Education 3	29,24	33,41
Private health insurance	86,12	90,69
No private health insurance	12,92	7,37
AMG	0,96	
CMU		1,94
Employed	61,21	67,83
Inactive	1,80	1,15
Homemaker	6,39	5,34
Retired	7,20	5,84
Unemployed	7,16	5,85
Student	16,23	13,98
Employee	27,69	26,98
Farmer	2,69	1,99
Self-employed	4,59	4,80
Executive	13,39	14,61
Technician	21,67	22,27
Skilled worker	19,18	21,07
Unskilled worker	10,02	7,67

Table 4.12: Descriptive statistics (*1998 & 2004 IRDES-HHIS*).

The mean value of equivalent income was lower in 1998[22], as well as the proportion of highly educated individuals. As for the employment status, the proportion of people in employment is lower in 1998 than in 2004: they represent 61% in 1998 and 68% in 2004. Over that period in France the number of unemployed people has slightly decreased while unemployment rates have markedly fallen: from 11,3% in December 1998 to 9.1% in December 2004, respectively from 13.84% to 10.10% for women, and from 10.20% to 7.80% for men[23]. Differences between actual proportions and our dataset are explained by both the age restricted sample and the inability of IRDES to interview precarious households. However our samples report lower proportions than actual observations from National statistics, descriptive statistics present similar trends. When it comes to the supplementary health insurance, the proportion of individuals without supplementary health insurance is lower in 2004, consequently, some of individuals might have enrolled CMU. As expected, the number of CMU beneficiaries represents a higher proportion in sample 2004 than the number of AMG beneficiaries in 1998 because CMU concerns a larger part of the population.

4.6.3 Explaining health within a linear model in 1998: comparisons with 2004

Prior to the comparative analysis, we shall measure income-related inequality in health in 1998. As we have previously done with data in year 2004, the measurement of income-related inequality in health follows a specific series of computations: health regression models, global concentration index of inequality in health and decomposition of this global concentration index.

This subsection deals with the first step. The table 4.13 presents results of interval regression models of the two mappings in 1998. Moreover, for the sake of comparison we have reported results for same mappings in 2004. In the first step, we briefly comment results in 1998 and point out differences between the two mappings. In the second step, we focus on changes in regression results over the two years.

It is noteworthy that the two mappings of the 11-categories self-assessed health in 1998 give similar results when significant. It is remarkable that mapping on SF6D leads to less significant effects than mapping on the health index. An explanation can be proposed with the distribution of SF6D (cf. figure 4.14) which varies less than the health index when self-assessed health is reported between 6 and 9.

Health decreases with age for both genders in 1998. Nevertheless, incidentally, there is no significant effect on health of women aged 36 to 45 years old for both mappings in 1998.

[22]We remind that we consider inflation-adjusted euros.

[23]INSEE `www.insee.fr`, La France en faits et en chiffres.

Variables	1998						2004					
	Predicted SF6D on 11-SAH			Predicted health index on 11-SAH			Predicted SF6D on 11-SAH			Predicted health index on 11-SAH		
	Coeff.	S.E	P-value	Coeff.	S.E	P-value	Coeff.	S.E	P-value	Coeff.	S.E	P-value
F 16-25	0,0306	0,0044	0	0,0358	0,0057	0	0,0290	0,0047	0	0,0344	0,0060	0
F 26-35	0,0208	0,0038	0	0,0221	0,0049	0	0,0060	0,0040	0,130	0,0085	0,0052	0,100
F 36-45	-0,0036	0,0037	0,332	-0,0047	0,0049	0,334	-0,0093	0,0039	0,016	-0,0119	0,0050	0,018
F 46-55	-0,0328	0,0038	0	-0,0468	0,0050	0	-0,0345	0,0039	0	-0,0485	0,0051	0
F 56-65	-0,0444	0,0049	0	-0,0659	0,0066	0	-0,0474	0,0049	0	-0,0695	0,0065	0
M 16-25	0,0546	0,0044	0	0,0585	0,0056	0	0,0489	0,0045	0	0,0516	0,0058	0
M 26-35	0,0309	0,0036	0	0,0300	0,0047	0	0,0149	0,0039	0	0,0160	0,0051	0,002
M 36-45	ref.											
M 46-55	-0,0234	0,0036	0	-0,0301	0,0048	0	-0,0247	0,0038	0	-0,0291	0,0050	0
M 56-65	-0,0439	0,0049	0	-0,0592	0,0066	0	-0,0414	0,0049	0	-0,0544	0,0065	0
Log income	0,0077	0,0017	0	0,0144	0,0022	0	0,0069	0,0017	0	0,0119	0,0023	0
Education less												
Education 2	0,0036	0,0023	0,124	0,0073	0,0030	0,016	0,0052	0,0026	0,042	0,0110	0,0033	0,001
Education 3	-0,0007	0,0025	0,769	0,0057	0,0032	0,073	0,0068	0,0026	0,009	0,0137	0,0033	0
Private health insurance	-0,0025	0,0027	0,346	0,0036	0,0034	0,289	0,0110	0,0035	0,002	0,0186	0,0045	0
No private health insurance	ref.											
CMU	.	.	.	.	.	.	-0,0068	0,0075	0,359	-0,0118	0,0098	0,228
AMG	0,0098	0,0089	0,269	0,0172	0,0114	0,132	.	.	.	.	.	.
Employed	ref.											
Inactive	-0,0789	0,0062	0	-0,1652	0,0084	0	-0,0360	0,0085	0	-0,0752	0,0111	0
Homemaker	-0,0044	0,0036	0,223	-0,0111	0,0048	0,021	-0,0040	0,0042	0,339	-0,0070	0,0055	0,202
Retired	-0,0108	0,0045	0,015	-0,0199	0,0060	0,001	-0,0034	0,0048	0,475	0,0042	0,0064	0,514
Unemployed	-0,0169	0,0034	0	-0,0258	0,0045	0	-0,0083	0,0040	0,038	-0,0097	0,0052	0,063
Student	-0,0001	0,0037	0,974	-0,0088	0,0046	0,058	0,0032	0,0040	0,434	-0,0053	0,0051	0,300
Employee	ref.											
Farmer	-0,0047	0,0054	0,384	-0,0090	0,0071	0,205	0,0109	0,0067	0,103	0,0182	0,0087	0,036
Self-employed	0,0171	0,0043	0	0,0208	0,0056	0	0,0136	0,0046	0,003	0,0138	0,0059	0,020
Executive	0,0167	0,0033	0	0,0150	0,0042	0	0,0078	0,0034	0,024	0,0097	0,0045	0,030
Technician	0,0059	0,0026	0,024	0,0071	0,0034	0,036	0,0082	0,0028	0,003	0,0117	0,0036	0,001
Skilled worker	0,0079	0,0027	0,003	0,0061	0,0035	0,081	0,0002	0,0029	0,937	-0,0012	0,0038	0,744
Unskilled worker	0,0005	0,0032	0,863	-0,0072	0,0041	0,081	-0,0029	0,0038	0,442	-0,0051	0,0049	0,294
Constant	0,6013	0,0118	0	0,8012	0,0154	0	0,5938	0,0131	0	0,7727	0,0171	0

Table 4.13: Health regression results in 1998 and in 2004(*1998 & 2004 IRDES-HHIS*).

The log of income significantly increases with health for both mappings. Retired, inactive and unemployed people have a negative and significant lower health status than employed people. For instance, inactive people have a health status that is 0.08 lower than those in employment when self-assessed is mapped on SF6D and 0.17 lower when self-assessed is mapped on the health index. Indeed, inactivity affects people having a poor health status in the restricted sample of 16-65 years old that we considered. As for socioeconomic status, it emphasises that self-employed people, executives, technicians and skilled workers have a significantly higher health status. Unskilled workers have a lower health status when significant. Our results show that insurance characteristics have no significant effects on health regardless of the health indicator. These results are interesting to compare with results in 2004, and we shall return to this point in the following comparison paragraph. Finally, despite the intuitive correlation that links education and income levels, education variables are non-significant. We cannot explain this surprising absence of significance, especially because other social variables exhibit expected relationships between health and socially advantaged groups.

By comparison to 2004, similar effects of age and gender categories on health are observed. Health decreases with age. Nevertheless, the age health gradient appears less strong in 2004. In the two years, income has a positive and significant effect on health which tends to be higher in 1998 than in 2004 for respective mappings of self-assessed health. When significant, the other social variables such as activity or social status describe similar effects in the two years: health increases with socially-advantaged characteristics. Analogously to income, their effects on health tend to be lower in 2004. The effect of education on health differs in the two years. There is a lack of significance in 1998 when health relies on the SF6D mapping and effects are reduced when the health index is used. The level of education has improved in France over the last ten years: more people are reaching university education and vocational training for adults has increased (OECD, 2007). In 2004, the proportion of people having a supplementary health insurance is lower than it is in 1998. Therefore, it appears that some people who have enrolled CMU were likely to ask for a supplemental health insurance before the reform. The lack of significance of AMG on health in 1998 shows that this policy was in fact different than CMU: first, it concerned less people and its granting was subject to unobserved parameters different from health status or income level, such as geographical area or health care utilisation.

4.6.4 Income-related inequality in health in 1998: comparisons with 2004

The second step of the analysis of inequalities in health evaluates the level of inequalities in health in 1998 using a concentration index of income-related inequalities in health. In table 4.14 we present the concentration indices for health in the two years for each mapping of the self-assessed health in eleven categories. Qualitatively, there is the same pattern over time for each mapping. In both years, an inequality in the distribution of health favouring the richest is observed in the French population. Nevertheless, these inequalities have changed over time and inequalities in health in 1998 were larger than in 2004.

Health indicators	1998			2004		
	CI	Newey-West S.E	95% Conf. Int.	CI	Newey-West S.E	95% Conf. Int.
Predicted SF6D on 11-SAH	0,0028	0,0003	[0,0013; 0,0025]	0,0020	0,0003	[0,0021; 0,0035]
Predicted health index on 11-SAH	0,0062	0,0003	[0,0037; 0,0049]	0,0043	0,0003	[0,0056; 0,0069]

Table 4.14: Comparisons between concentration indices for the income-related health inequality (*1998 & 2004 IRDES-HHIS*).

The magnitude of inequality differs with the mapping and the health index describes a stronger inequality in each year. However, these differences in magnitude between the two mappings and changes in the level of inequalities in health are similar and proportionate to the differences in mappings[24]. From this global result, it is interesting to understand the mechanisms underlying these changes. In particular, we would like to know which individual characteristics have played a role in the reduction of income-related inequalities in health over that period. The decomposition of these inequalities will allow us to answer this question.

4.6.5 Decomposition of inequalities in 1998: comparisons with 2004

The last step of the analysis of inequalities in health is the decomposition of the income-related inequality in health into the contribution of each of individual characteristics to the inequality. We replicate the decomposition method to year 1998. Table 4.15 describes the contribution of each regressor to the global concentration index of inequality in health in 1998. In order to facilitate comparisons, we have reported results presenting the same mapping of health in 2004. The last three rows of table 4.15 recapitulates the income-related inequality in health and gives the level of inequity in health. Inequities in health favouring the richest individuals are observed. Analogously to the global inequality in health, the inequality in health is larger when self-assessed health is mapped on the health index in both years and a reduction of its level is observed over the period.

Qualitatively, concentration indices emphasise similar pattern over time. Contributions to inequality of regressors move in parallel to the regression coefficients that we have previously described. The analysis of changes in concentration indices as well as in elasticities allows us to understand the reduction of income-related inequalities in health between 1998 and 2004.

[24]When self-assessed health is mapped on SF6D then $|\frac{CI_{2004}-CI_{1998}}{CI_{2004}}| = 0.4$ and when self-assessed health is mapped on the health index then $|\frac{CI_{2004}-CI_{1998}}{CI_{2004}}| = 0.44$.

It is remarkable that the most important contribution to income-related inequalities in health stem from the log of income in both years. Income stays the most important contributor in both years. The income concentration index is weakly lower in 2004 so we cannot say that changes between the two years are explained by a reduction in inequalities in income. Nevertheless, the health elasticity of income decreases between the two years and changes are thus explained by changes in the strength of the association between income and health. It appears also that income explains a larger proportion of inequalities in health in 2004 than in 1998. Indeed, income-related inequalities in health in 1998 are explained by various activity status and social status characteristics whose contributions are less pronounced in 2004. For example, being an executive drives a higher inequality over income in 1998 and also contributes more to a pro-rich distribution of health. Similarly, we notice changes into concentration indices of unemployed and inactive people, who suffer from a higher inequality over income in 1998.

In both years, farmers and unskilled workers experience the strongest inequality in income whereas executives and technicians are concentrated in higher income groups. It is interesting to underline that it is the concentration index associated to farmers that has changed the most between 1998 and 2004. Considering that the elasticity of health with this social status has not changed, this reduction is mainly driven by changes in the distribution of income among farmers.

Whereas the second most important contributor of the income-related inequalities in health in 2004 is education, its contribution to inequalities in 1998 is very weak. Moreover, in 1998 the contribution of a higher education is even unexpectedly negative when self-assessed health is mapped on SF6D. This negative contribution comes from the negative elasticity of health with higher education. Consequences of this unexpected negative sign are negligible as the contribution of education to the explanation of inequality in health in 1998 is weak. As for the mapping on the health index in 1998, its positive contribution to inequalities is also weak. As a result, differences with results in 2004 are in line with the lack of significance of the regression coefficient of education in 1998. The positive effect of a higher education in 2004 is due to the fact that better educated people are more efficient producers of health because they have a better knowledge of healthy behaviours (Grossman, 1972).

Age and sex categories are important contributors in both years. Younger age classes have similar negative contributions to inequalities in 1998 and 2004 because of their concentration in lower income levels despite their good health status.

Variables	CI		1998						2004					
			Predicted SF6D on 11-SAH			Predicted health index index on 11-SAH			Predicted SF6D on 11-SAH			Predicted health index index on 11-SAH		
	1998	2004	Elast.	Contrib.	%	Elast.	Contrib.	%	Elast.	Contrib.	%	Elast.	Contrib.	%
F 16-25	-0,199	-0,173	0,005	-0,0010	-36,62%	0,004	-0,0009	-14,11%	0,005	-0,0008	-42,14%	0,004	-0,0007	-16,92%
F 26-35	0,016	-0,020	0,004	0,0001	2,13%	0,003	0,0000	0,75%	0,001	0,0000	-1,02%	0,001	0,0000	-0,49%
F 36-45	-0,056	-0,078	-0,001	0,0000	1,23%	-0,001	0,0000	0,53%	-0,002	0,0001	6,47%	-0,002	0,0001	2,82%
F 46-55	0,123	0,094	-0,005	-0,0006	-21,03%	-0,005	-0,0006	-9,91%	-0,006	-0,0005	-28,03%	-0,006	-0,0006	-13,32%
F 56-65	0,081	0,188	-0,004	-0,0003	-11,77%	-0,004	-0,0004	-5,77%	-0,005	-0,0009	-47,23%	-0,005	-0,0010	-23,39%
M 16-25	-0,132	-0,141	0,010	-0,0013	-47,73%	0,008	-0,0010	-16,89%	0,009	-0,0012	-60,80%	0,007	-0,0009	-21,68%
M 26-35	0,055	0,020	0,005	0,0003	9,93%	0,004	0,0002	3,17%	0,002	0,0000	2,25%	0,002	0,0000	0,82%
M 36-45														
M 46-55	0,137	0,120	-0,004	-0,0005	-17,33%	-0,003	-0,0005	-7,36%	-0,004	-0,0004	-22,84%	-0,003	-0,0004	-9,11%
M 56-65	0,105	0,186	-0,004	-0,0004	-15,39%	-0,004	-0,0004	-6,85%	-0,004	-0,0007	-36,97%	-0,004	-0,0007	-16,41%
Log income	0,047	0,041	0,083	0,0039	137,98%	0,113	0,0052	84,99%	0,078	0,0032	163,94%	0,100	0,0041	95,44%
Education less														
Education 2	0,015	-0,032	0,001	0,0000	0,65%	0,002	0,0000	0,44%	0,002	-0,0001	-2,74%	0,003	-0,0001	-1,96%
Education 3	0,276	0,242	-0,0003	-0,0001	-3,12%	0,002	0,0005	8,18%	0,003	0,0008	42,71%	0,005	0,0013	29,10%
Private health insurance	0,058	0,041	0,003	0,0002	6,71%	0,004	0,0002	3,22%	0,015	0,0006	31,41%	0,019	0,0008	18,06%
No private health insurance														
AMG	-0,695	.	0,0001	-0,0001	-3,54%	0,0002	-0,0001	-2,05%	.	.	.	.	.	.
CMU	.	-0,711	.	.	.	.	.	.	-0,0002	0,0001	6,75%	-0,0002	0,0002	3,93%
Employed														
Inactive	-0,372	-0,303	-0,002	0,0008	28,44%	-0,003	0,0012	19,65%	-0,001	0,0002	9,80%	-0,001	0,0003	6,91%
Homemaker	-0,254	-0,327	-0,0004	0,0001	3,85%	-0,001	0,0002	3,19%	-0,0003	0,0001	5,47%	-0,0004	0,0001	3,24%
Retired	0,112	0,184	-0,001	-0,0001	-4,67%	-0,002	-0,0002	-2,84%	-0,0003	-0,0001	-2,87%	0,0003	0,0001	1,18%
Unemployed	-0,363	-0,294	-0,002	0,0007	23,60%	-0,002	0,0007	11,88%	-0,001	0,0002	11,05%	-0,001	0,0002	4,38%
Student	-0,141	-0,170	0,000	0,0000	0,15%	-0,002	0,0002	3,57%	0,001	-0,0001	-5,86%	-0,001	0,0001	3,31%
Employee														
Farmer	-0,408	-0,285	-0,0002	0,0001	2,74%	-0,0003	0,0001	1,75%	0,0003	-0,0001	-4,81%	0,0004	-0,0001	-2,72%
Self-employed	-0,079	-0,034	0,001	-0,0001	-3,34%	0,001	-0,0001	-1,34%	0,001	0,0000	-1,75%	0,001	0,0000	-0,60%
Executive	0,522	0,471	0,003	0,0018	62,86%	0,002	0,0011	18,64%	0,002	0,0008	41,85%	0,002	0,0008	17,58%
Technician	0,242	0,211	0,002	0,0005	16,51%	0,002	0,0004	6,61%	0,003	0,0006	29,86%	0,003	0,0006	14,50%
Skilled worker	-0,216	-0,222	0,002	-0,0005	-17,67%	0,001	-0,0003	-4,51%	0,0001	0,0000	-0,84%	-0,0003	0,0001	1,51%
Unskilled worker	-0,394	-0,366	0,0001	0,0000	-1,16%	-0,001	0,0003	5,05%	-0,0003	0,0001	6,33%	-0,0004	0,0002	3,80%
Total CI				0,0028			0,0062			0,0019			0,0043	
CI*				-0,0038			-0,0035			-0,0045			-0,0042	
I=CI-CI*				0,0067			0,0096			0,0064			0,0085	

Table 4.15: Decomposition of inequalities in health in 1998 and in 2004 (*1998 & 2004 IRDES-HHIS*).

On the contrary, there are relevant changes in older age categories. The contribution to inequalities of people aged 56 to 65 years old, especially women, is stronger in 2004 than it was in 1998. This change is driven by higher concentration indices for these age categories in 2004, people aged 56-65 being more concentrated in higher income levels in 2004 than they were in 1998. Analogously, retired people are more concentrated in higher incomes in 2004 than they were in 1998.

Finally, we shall consider changes in insurance coverage. In both years, having a supplementary health insurance is concentrated in higher incomes and displays a positive contribution to inequalities. These effects are increasingly important in 2004. Between the two years, the elasticity of health with private health insurance has increased. In 2004, people are more likely to ask for a private health insurance when their health status is poor. This finding is line with the literature pointing out the existence of health-related choices in insurance (Francesconi *et al.*, 2006; Couffinhal, 1999).

As for AMG in 1998 and CMU in 2004, beneficiaries of these reforms are heavily concentrated in the poorest income levels. Nevertheless, they contribute differently to the income-related inequalities in health. Indeed, CMU contributes to an increase of inequalities because of the negative elasticity of health with CMU. This finding confirms that the beneficiaries of CMU are relatively in worse health than others. The reform concerns thus people who need it. On the contrary, the elasticity of health with AMG is negative. Even if AMG concerns the poorest people, it appears that AMG is not related to health status as previously shown with regression coefficients. These results confirm that the contribution of CMU to income-related inequalities is not negligible.

4.7 Conclusion

As compared to the existing literature in France, this analysis is relevant for several reasons. Firstly it uses more recent data than the existing literature. Secondly, it is not restricted to a cross section approach for one survey year but investigates a comparative of social health inequalities over the last decade. Thirdly, the stochastic dominance analysis completed by the decomposition method allows a better understanding of the elements involved in the existence of inequalities at both static and dynamic levels. Fourthly, it uses innovative measurements of health. Finally, it achieves a more reliable measurement of inequality due to the use of an interval regression approach to estimate a fully specified health equation.

The analysis of income-related inequalities in health shows that France experiences inequalities in health to the detriment of the poorest. These results are qualitatively analogous to those reported in an European study involving France data from the 1996 ECHP[25]

[25] European Community Household Panel

(van Doorslaer & Koolman, 2004), which studies inequalities in health in the whole population from 16 years old. The decomposition of inequality in health in 2004 shows that a higher income, a higher education level and a higher socioeconomic status, such as executive or technician are strongly contributing to income-related inequality in health. Therefore, income does not act on health in isolation from other factors. Indeed, education as well as socioeconomic status are other important factors that influences health. These results for education and income coincide with the European results comparisons (van Doorslaer & Koolman, 2004). As for CMU, this reform was proposed in 2000 and its positive contribution to inequality in health relies on the fact that poor people in very bad health are more likely to ask for a free health care coverage. Our analysis of inequalities confirms that the reform concerns the targeted population but the time period is too short to observe global changes on health status. The strong contribution of income to the inequalities in health emphasises that measures which can reduce either the health-harming effects of income losses or the income consequences of health losses could reduced inequalities in health (van Doorslaer & Koolman, 2004).

The over time analysis has emphasised some changes in income-related inequalities in health in France. It appears that social inequalities in health have decreased, driven by a lower elasticity of health with income and by lower inequalities over income of some specific social groups such as inactive, unemployed and even executives. The strength of the association between health and private health insurance has also changed over time and in 2004, people are more likely to ask for a private health insurance when their health status is poor. It appears also that individuals aged 56-65 years old are socially less disadvantaged in 2004.

These results need to be confirmed on larger samples and longitudinal data. Indeed, analyses with longitudinal data show that there are important features of income-related inequalities in health that cannot be revealed by cross-sectional data (Jones & Lopez, 2004). Similarly, better information of income might increase the relevance of our analysis, which also suffers from the inability of IRDES-HHIS to interview the poorest households. Nevertheless, this result is very original as regard to other French studies which show that socioeconomic inequalities in mortality or specific diseases are increasing (Leclerc *et al.*, 2000; de Koninck & Fassin, 2004).

Another considerable contribution of this chapter is to involve sophisticated health indicators suitable for the measurement of inequalities in health in France. Firstly, the use of the thresholds of HUI in the French context is particularly relevant as the van Doorslaer and Jones (2003) mapping as turned out to be the preferred tool in the most recent European studies of social inequalities in health. Secondly, the use of the SF6D utility algorithm to estimate a preference-based measure of health from the French SF36 has no precedent. It allows researchers to use specific econometric models, such as interval regression, and it might increase the number of uses that could be done from the French SF36 questionnaire in econometric analyses and economic evaluation studies. Finally, the health index generated in chapter 2 is empirically tested. So far, it was unclear to what extent social inequalities in health in France were sensitive to recent measurement of health which rely promising construction methodologies and do not concern ill-health. We have discussed the influence of the measurement of health on inequalities in health at two relevant levels: firstly, as regard to the distribution of health used in the mapping, and secondly as regard to the scale of self-assessed health. It appears that the magnitude of income-related inequalities in health is sensitive to the spreading of the distribution of health. For example, when a distribution is concentrated such as HUI in good health statuses, it always induces a lower level of inequalities. Conversely, the distribution of the health index offers a larger standard deviation in health status from poor to very good, which increases differences and *a fortiori* inequalities. The income-related inequalities in health are also sensitive to the number of categories of self-assessed health. A lower number of categories is likely to perceive less distinctions among health statuses and imply higher concentration indices, as shown by the lower concentration indices observed for the mapping of the 11-categories self-assessed health with both the health index and SF6D. The choice of the health distribution to scale self-assessed health therefore has consequences. In this connection, we find that the health index is a valid indicator for the study. Indeed, it qualitatively displays similar patterns as mapping on SF6D, which is presumed as the "gold" health indicator. The relevant similarities from a qualitative point of view with other mappings such as SF6D or HUI confirm its validity to measure health status. Moreover, the distribution of the health index presents an advantage in comparison with the two other distribution of health because it describes health from 0 to 1. Indeed, the health utility index as well as the SF6D have no natural zero point[26].

Chapter 4 offers also to use an appropriate econometric modeling. the thresholds used in the interval regression can be allowed to be different for different groups of individuals or when comparing across different countries as they depend on the relative frequencies in each category of self-assessed health. Moreover, as the thresholds determine the scale of the latent variable; this is equivalent to allowing for heteroscedasticity in the error term of the latent variable specification.

To conclude, we shall underline that inequalities in health have been specifically analysed here as regard to individual's current conditions, ignoring that these current conditions, as well as the health status, might be strongly related to individual's past conditions.

[26]Some Canadian surveys after the wave 1994, include negative health utility index scores. It would mean that there are health statuses worse than death. In this context, the health utility index cannot be compared to a ratio-scale variable and the main assumption formulated by van Doorslaer and Jones (2003) would not be respected.

Consequently, chapter 5 will propose an analysis of inequalities in health enlarged to other determinants of health.

Chapter 5

From inequalities in health to inequalities of opportunity in health

This chapter is part of the research program "Inégalités sociales de santé", supported by DREES-MiRe, Inserm, DGS, InVS, INCa and CANAM. Some of the analyses are a joined work with Marion Devaux, Florence Jusot (IRDES) and Alain Trannoy (IDEP). Primary results of this study have been published in Issues in Health Economics Series[1] *and in the proceedings of the 27th AES Conference*[2].

5.1 Introduction

Many European studies show strong and long-lasting inequalities in health related to current socioeconomic status (Wagstaff & van Doorslaer 2000b; Mackenbach *et al.*, 1997). These inequalities in health have been extensively explained by differences in health status due to living and working conditions, access to health care, or health-related behaviours such as beneficial or risky choices on French data (Leclerc *et al.*, 2000; Monteil *et al.*, 2005), as well as on data of other European countries (van Doorslaer & Koolman, 2004; Simon *et al.*, 2000). Analogously, in chapter 4, an explanation of inequalities in health according to current individual characteristics has been carried out and has shown that income, social status and activity status as well as health insurance coverage are important contributors to these income-related inequalities in health. Nevertheless, as we underlined

[1] Devaux M., Jusot F., Trannoy A. and Tubeuf, S. (2007). Inégalités des chances en santé: influence de la profession et de l'état de santé des parents, Questions d'économie de la santé n^{o}118. Série Résultats IRDES.

[2] Devaux M., Jusot F., Trannoy A. and Tubeuf, S. (2007). Origine sociale et état de santé des parents: Quelle influence sur l'état de santé à l'âge adulte ? in "Approches institutionnalistes des inégalités en économie sociale", Actes des XXVIIe Journées de l'Association d'Economie Sociale, Nanterre, 6 et 7 Septembre 2007.

in the introduction of this thesis, health defines itself over the whole life cycle and is not only influenced by current life conditions but also by other factors such as health behaviours or living conditions ten, twenty or even fifty years earlier. Indeed, some recent analyses, mainly epidemiological, document significant correlations between adult life and living conditions in childhood, even *in utero* (Smith, 1999; Goldberg *et al.*, 2002; Marmot & Wilkinson, 1999; Wadsworth, 1999; Power *et al.*, 1998). In particular, social background influences health status at three levels.

Firstly, social background influences **health status in childhood**. Malnutrition adversely affects not only bodily growth, but also cognitive development and educational attainment. Conversely, breast feeding of three months or more seems to be associated with improved cognitive performance (Wadsworth, 1999).

Secondly, social background influences **health status over the whole life cycle** by an accumulation of vulnerability during childhood and adult life. The 1958 British cohort has allowed Bartley *et al.* (1994) to give evidence to this accumulation process. Their analyses have found that babies with low birth weight, who have suffered of poor nutrition, are more likely to experience growth problems. In addition, they have a disadvantaged life trajectory and their health status in adulthood is affected by the accumulation of social and biological risks.

Thirdly, social background influences **health status in adulthood** in two ways: a direct way and an indirect one.

- Poorer conditions in childhood *directly* influence health status in adulthood following a latency period; it is the latency model (Barker, 1996; Wadsworth, 1999). There is a strong relationship between childhood conditions and health in adulthood, which can be compared to a biological programming. During childhood, a specific risk is established and it needs a trigger in adulthood to be reactivated. For example, health problems occurring in mid adult life, such as respiratory functions, may be more rapidly critical if maximal foetal and early childhood growth potential has not been achieved (Dezateux & Stocks, 1998). Barker (1997) shows that a high-risk of non-insulin-dependent diabetes and impaired glucose tolerance is found in people who experienced foetal malnutrition.

- Conditions in childhood, particularly parents' socioeconomic status *indirectly* influence health status in adulthood through a transmission of socioeconomic status (SES) over different generations (Case *et al.*, 2005). This second way is called the pathway model. In France, some studies using either the GAZEL cohort of employees from the national electricity and gas company (Hyde *et al.*, 2006; Melchior *et al.*, 2006a) or the Life History Survey (Melchior *et al.*, 2006b), have shown an indirect influence of the father's social status on both the health status and risk of death of their descendants. Power and Hertzman (1997) argue that adult disease is more fully understood when account is taken of the combined effects of social and biological risk in early life.

Our analysis focuses on the influence of social background on health in adulthood. In addition to the two previous ways of explanation, we argue that this influence could also be due to another characteristic seldom considered: parents' health status. Indeed, if on one hand, there are social inequalities in health in the parents' generation, and on the other hand, parental health status is correlated to their descendant's health status, then we can conclude that social background influences descendant's health status, although this influence would not imply a causality link. We call this third way of explanation *hypothesis of health transmission*. This hypothesis considers the influence of parents' health status on their descendant's health status and relies on four elements that could be transmitted from one generation to another.

- Firstly, it relies on health capital models (Grossman, 1972). According to these models, health is likened to a capital, which evolves over time according to age, individual health behaviours and investments in health, and stays strongly influenced by its initial level. This initial level is partly related to parents' health status through a common genetic inheritance.
- Secondly, parents take into account their own health status in the family decision of investment in the health capital of each family members and then in their child health capital (Jacobson, 2000; Bolin *et al.*, 2001; Bolin *et al.*, 2002).
- Thirdly, lifestyles in adulthood, such as nutrition, exercise, smoking and alcohol use rely on an effective imitation of parents' behaviours. According to McLeroy *et al.*, (1988), health-related behaviours are determined by five categories of factors: interpersonal factors, interpersonal processes and primary groups, institutional factors, community factors and public policy. Therefore, two of these five factors concern parents. Their health-related behaviours can thus influence their offspring's lifestyle, from childhood to adulthood (Fulghum, 1986).

  > "Don't worry that children never listen to you; worry that they are always watching you".

- Finally, we can also assume that parents pass on their preferences for health to their children. Individuals in adulthood may behave similarly to their parents for prevention choice and health care utilisation.

Whereas this transmission of health status has been shown for some specific health diseases, such as cardiovascular diseases (Blane *et al.*, 1996; Poulton *et al.*, 2002), the transmission of general health status in adulthood has not been considered at present. This third way is reinforced by recent analyses, which confirm the influence of parents' health status on their children's health status (Case *et al.*, 2002; Llena-Nozal, 2007). Moreover the persistency of this effect on a descendant's health over the whole life-cycle, especially in adulthood, has never been studied because of a lack of data gathering together with individual general health information and parents' health in most health surveys. This research aims to fill this gap, examining the intergenerational transmission of inequalities in health using a survey carried out on a general population.

Furthermore, this research context offers an empirical work on inequalities of opportunity in health, which are a central line of research in the process of development (World Bank, 2005). Although inequalities of opportunity in health are implicitly analysed within equity in health and in health care utilisation studies (Gravelle, 2003; van Doorslaer *et al.*, 2000; van Doorslaer *et al.*, 2002), Dias and Jones (2007) have recently drawn attention to the fact that equality of opportunity should be given a fair innings in health economics. As both social background and parents' health represent circumstances independent of individual responsibility (Dworkin, 1981; Arneson, 1989; Roemer, 1998), the distribution of health in adulthood conditional on these circumstances will describe inequalities of opportunities in health.

This analysis of inequalities of opportunity in health related to a family and social determinism in France[3] can be compared to other studies that evaluate inequalities of opportunities in various spheres such as education, employment, housing or income distribution (Lefranc *et al.*, 2004).

In the first section, this chapter defines precisely the concept of inequalities of opportunity in the specific context of health. The second section describes data coming from the Survey on Health, Ageing and Retirement in Europe (SHARE) and the indicators involved in the study, such as parents' health, which is measured by their relative longevity in comparison with their birth cohort. The measurement of inequalities of opportunities in health follows both methods from analyses of equality of opportunity in income and typical explaining health studies. The first step displayed in the third section, consists in stochastic dominance analyses as presented in chapter 3. The second step relies on a parametric approach, which explains health status in adulthood according to family and social background as well as current social characteristics. It is presented in the fourth section. Finally, in the fifth section, we propose an original use of concentration indices to measure inequalities of opportunity in health according to parents' health status. Discussion and concluding remarks on these inequalities of opportunity in health form the last section.

[3] A report (Boarini *et al.*, 2006) treats this question in the context of a theoretical and empirical analysis of social justice norms in terms of health in different European countries. Unlike our analysis, this research does not give any evidence of inequalities of opportunity in health in France.

5.2 Equality of opportunity in health

The intergenerational equality of opportunity in health is reached when comparing different cumulative distribution functions of health status across several sub-groups of individuals and which are distinguished by a characteristic of the parental generation.

Health statuses are more generally described with qualitative than quantitative variables. The distribution of health status over different categories allows definition of the proportion of individuals with a given health status in the same generation. One can then easily represent the cumulative distribution function of health statuses. It shows the proportion of individuals within a specific health status, which is at least equal to a given category. For instance, individuals with poor health would represent 30%, those with poor and fair health 60% etc. This cumulative distribution function of the general population can be interpreted as a distribution of opportunities. Indeed, a randomly selected individual has a 30% chance of belonging to the category in poor health. To liken this distribution to a draw of lottery tickets is improper because an individual actively contributes to the outcome. In the context of health, one can actively contribute to improve or to worsen it.

Consider now that instead of a simple description of health statuses over a population, one is interested in the distribution of health statuses according to specific characteristics of childhood conditions. For instance, we graph the cumulative distribution function of the health status of individuals born either of a father who was a blue collar worker or of a prematurely dead father. Being the son of a blue collar worker as well as being the son of a prematurely dead father is obviously an exogenous characteristic; descendants have no control over these factors. The fact of being born in a particular family background is equivalent to get a lottery ticket, whose winnings will only be known later on. The cumulative distribution function of health status of individuals born to blue collar workers, 30, 40, or 50 years later describes the distribution of equality of opportunities in the health of sons of blue collar workers. If on one hand, this cumulative distribution function is clearly different than the one of individuals born to white collar workers and if on the other hand, this difference is such that a descendant has a higher chance of being in poor health when he is born to a blue collar worker, one can reasonably associate this result to a difference in social backgrounds. The previous example is a typical situation of stochastic dominance at first order. Graphically, the cumulative distribution function of health statuses of individuals born to a blue collar worker is always above that of individuals born to a white collar worker at any point of comparison. In this context, the comparison of random distributions of health statuses conditional on family background leads any individual to prefer systematically being born to a white collar worker than born to a blue collar worker regardless of his risk-aversion. There is thus a social inequality of opportunity in health. Conversely, if two cumulative distribution functions are the same, then one concludes that there is a social equality of opportunity in health. The equality of opportunity is equivalent to a situation where an individual would be indifferent to the choice of a family background.

The same approach can be proposed when comparing sub-groups of individuals according to parents' health. Analogously, if there are no differences between two cumulative distribution functions of health status, one would conclude on an equality of opportunity in health. However, in this context, this equality cannot be called "social" because of the intergenerational transmission of genes but one would conclude there is a "health" inequality of opportunity in health. An interpretation of a "complete" intergenerational equality of opportunity in health would superficially be that family background does not endow any advantages not only on average but also on any percentile of the distribution of health statuses. As a result, either one is in poor health or good health; social background is not a determinant of health status. In this context, if the descendant is in poor health, this could be explained either by risky health behaviours or misfortune in adult life. The distribution of health status is the result of misfortune and factors within the control of the descendant.

Empirically, the inference procedure relies on tests of stochastic dominance at first order as described in Lefranc *et al.* (2004) and in appendix A. As distributions considered in this study are discrete, our approach will be limited to unilateral *Kolmogorov-Smirnov* tests of equality of distribution.

One of the difficulties of our dominance analysis is that it assumes the availability of large samples. If we intersect every possible social background with other different criteria, then sample size reduces and the dominance tools cannot be used any longer. Consequently, a multivariate regression analysis is then proposed in order to supplement the dominance analysis. The dependant variable of this parametric analysis is the descendant's self-assessed health. We would like to underline that an analyst of inequality of opportunity is first interested by the potential correlation between family background and descendant's self-assessed health from an ethical point of view, whereas an econometrician will initially be looking for causality links. The correlation we study gives intuitions on the causality link since an individual has no control on his family background. The regression analysis offers flexibility to test for a variety of hypotheses that could not be considered in the dominance approach; however, a parametric context is always more restrictive than a dominance approach, which is essentially non-parametric.

5.3 The French part of SHARE: a relevant tool for empirical work

The different hypotheses on the relationship between family background and health in adulthood have never been tested on French data at general population level because of a lack for suitable data. Despite numbers of French health surveys, none of them gather both childhood information and health in adulthood.

5.3.1 Data and sample

This study relies on the French part of the European survey called SHARE. As ageing is a social and economic challenge of the 21^{st} century in Europe, a panel on health, ageing and retirement in Europe was launched in 2004/2005 to study this phenomena with relevant data (Börsch-Supan *et al.*, 2003). It benefits from the experience of the American Health and Retirement Survey[4] and the English Longitudinal Survey of Ageing[5], and studies Europeans aged 49 years and older as well as their spouses.

Eleven countries[6] are involved in the project and the common set-up of data is strictly comparable in order to facilitate cross-country comparisons. Different disciplines are covered by the questionnaire namely demography, economics, sociology, and epidemiology. In addition to questions about their current situation, individuals are asked about past circumstances, in particular about parents. Although SHARE will be a longitudinal collection of data, at present only the first wave of the survey is available. Furthermore, for technical reasons we only have access to the socioeconomic status codification for French data[7]. However, a European comparison of this question would be interesting to consider as soon as data permit it.

For the first time in France, this survey permits linking an individual's health status in adulthood with his social background on a representative sample. Both parents' final social status and demographic characteristics (age at death for deceased parents and age at the time of the survey for parents still alive) are available.

Considering the respondents' age (49 and over), a large proportion of them have lost their parents (84% of the initial SHARE sample). In order to test the influence of social background and parents' health on health status in adulthood, the analysis is focused on individuals whose parents have died when the survey was conducted and answered

[4] Detailed information on AHRS can be found on `http://hrsonline.isr.umich.edu`.

[5] Detailed information on ELSA can be found on `www.ifs.org.uk/elsa`.

[6] Austria, Belgium, Denmark, France, Germany, Greece, Italy, The Netherlands, Spain, Sweden and Switzerland.

[7] As the French National Institute of Statistics (INSEE) and IRDES are taking part to the survey organisation, our access to the French part of the survey has been anticipated. However, this access has been restricted to a limited number of variables. This limitation of data will be particularly important in the empirical reasoning as it restricts the number of instrument variables available.

questions about their self-assessed health, last job or occupation, their parents' final job or occupation and age at death. The analysis sample is composed of 1783 individuals.

5.3.2 Variables measuring social conditions

Social background

In SHARE, social background is measured by the last job or occupation the father or mother of the respondent had. The ISCO classification (International Standard Classification of Occupations) is used for categorising occupations. Jobs are classified with respect to the type of work performed. The basic criteria used to define the system of major, sub-major, minor and unit groups are "skill level" and "skill specialisation" , which are required to carry out the tasks and duties of the occupations. This classification emphasises ten main groups of occupation (Elias, 1997). In this analysis, people have been gathered into six socioeconomic status groups for the fathers[8]. As shown on the figure 5.1, more than one third of respondents are born to a craftsman or a skilled worker (35%) and 27% of respondents' fathers worked in farming. About 13% of fathers belonged to a higher social class, i.e. were managers or professionals, the other professional groups are represented by proportions under 10%.

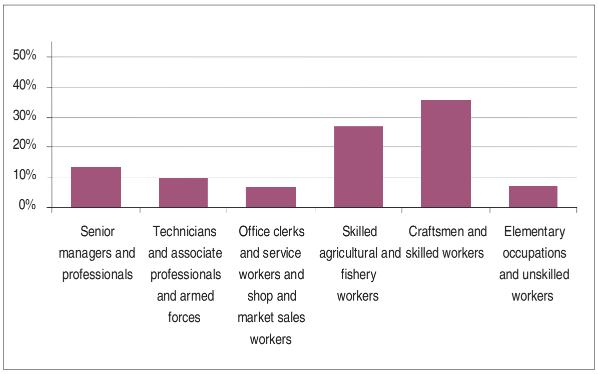

Figure 5.1: Distribution of the father's socioeconomic status (*2004/05 SHARE*)

[8]These six groups are (i) "senior managers and professionals", which is composed of "legislators, senior officials, corporate managers and managers of small firms", and "physical, mathematical, engineering science, life science, health, and teaching professionals"; (ii) "technicians and associate professionals" and "armed forces"; (iii) "office clerks" and "service workers and shop and market sales workers"; (iv) "skilled agricultural and fishery workers"; (v) "craftsmen and skilled workers" represents "craft and related trades workers" and "plant and machine operators and assemblers" and (vi) "elementary occupations and unskilled workers".

A classification of six groups[9] is proposed for mothers. As regard to the age of respondents, the group of mothers who were homemakers is represented by almost one half of the respondents (cf. figure 5.2). However, when the mother was active, she was mainly working in farming or belonged to elementary occupations.

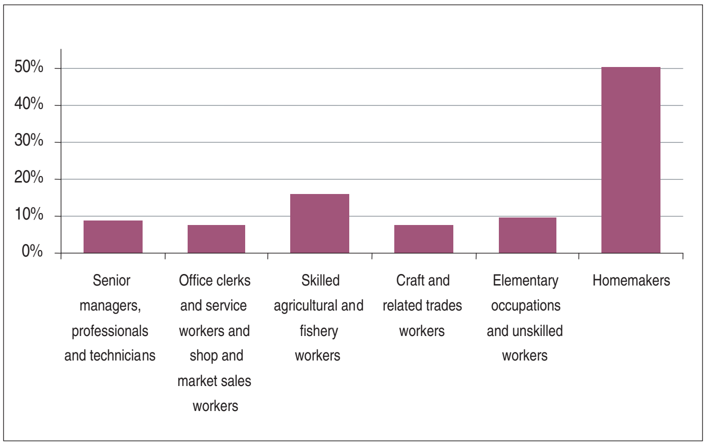

Figure 5.2: Distribution of the mother's socioeconomic status (*2004/05 SHARE*)

Current socioeconomic status of the descendant

Each respondent's current socioeconomic status is considered on two levels: education and social status. Education level is considered first and is measured by the highest diploma gained. In this way, education is described in four categories[10]. Then, current or last job is classified by ISCO into seven groups[11]. Among these respondents aged 49 years and older, the most important group is "office clerks and service and shop workers" (22%), then follows "craftsmen and skilled workers" representing 19%. "Senior managers and professionals" and "technicians and associate professionals, armed forces" respectively represent 17% and 16% of the sample whereas "elementary occupations and unskilled workers" equal 11% (cf. figure 5.3).

[9]The first group "senior managers, professionals and technicians", is composed of "legislators, senior officials, corporate managers and managers of small companies", "physical, mathematical, engineering science, life science, health, and teaching professionals", and "technicians and associate professionals". The second group contains "office clerks" and "service workers and shop and market sales workers"; the third "skilled agricultural and fishery workers"; the fourth "craft and related trades workers" and "plant and machine operators and assemblers"; the fifth "elementary occupations and unskilled workers" and the last one mothers, who were "homemakers".

[10]The four categories for education are: no diploma (23%), elementary level diploma (31%), secondary level diploma (27%) and baccalauréat (A-levels) (19%).

[11]These seven groups are (i) "senior managers and professionals"; (ii) "technicians and associate professionals" and "armed forces"; (iii) "office clerks" and "service workers and shop and market sales workers"; (iv) "skilled agricultural and fishery workers" ; (v) "craftsmen and skilled workers" and "plant and machine operators and assemblers" ; (vi) "elementary occupations and unskilled workers"; and (vii) "homemakers" .

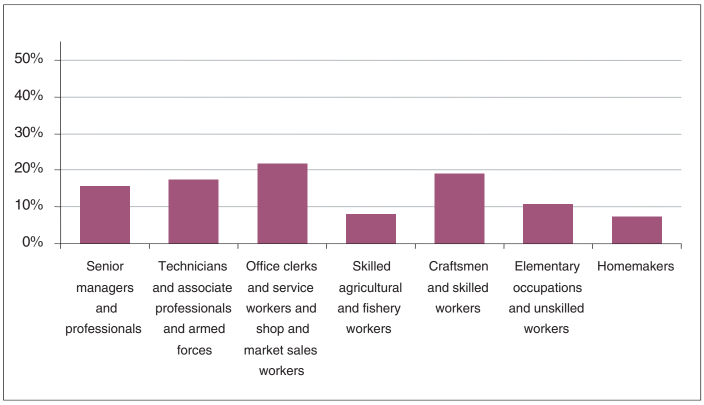

Figure 5.3: Distribution of the respondent's socioeconomic status (*2004/05 SHARE*)

5.3.3 Variables measuring health conditions

Health is a multidimensional parameter and is therefore difficult to summarise as a unique indicator. Generally, two types of indicators are used in health analyses: mortality and morbidity indicators. Our study relies on both types of indicators. Descendant's health status is measured by a morbidity indicator, self-assessed health, whereas parents' health is based on a mortality indicator, namely their relative longevity by comparison to their birth cohort.

Self-assessed health of the descendant

Self-assessed health is the most collected variable in European surveys on health, which are based on interview (Barnay *et al.*, 2005). Despite its subjectiveness, this indicator has been found to be a good health indicator, which predicts mortality (Idler & Benyamini, 1997) as well as health care utilisation (DeSalvo *et al.*, 2005). SHARE contains two questions on self-assessed health; the one, promoted by the RAND Corporation[12] and the one recommended by the European WHO (1996). They both rely on the same question: *"Would you say your health is ..."* but vary in response choices, respectively: *"excellent, very good, good, acceptable, poor"* and *"very good, good, fair, poor, very poor"*. In the following analysis, we consider the European wording for measuring respondents' health in adulthood. More than one half of respondents report a good health status: 45% report a good health status and 11% a very good one (cf. figure 5.4).

[12]Reports from the RAND Corporation can be found on `www.rand.org`.

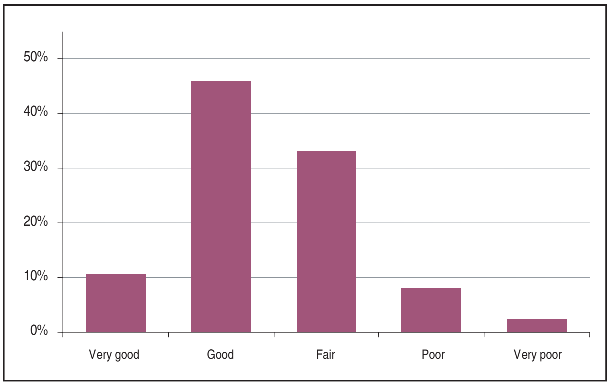

Figure 5.4: Distribution of the respondent's self-assessed health (*2004/05 SHARE*)

The measure of parents' health: the relative longevity

In general, it is difficult to construct a health indicator, which would be the same for the oldest and the youngest adults as regard to the depreciation of health with age. As we cannot use the usual self-assessed health to measure parents' health status, we have to construct a proxy of their health status. An old saying "*long life runs in families*" is supported by a large number of studies[13]. Longevity seems to be a relevant health measure for our analysis as well. This health indicator relies on parents' relative longevity compared to their expected longevity at birth. That is, the indicator equals the difference between age at death and life expectancy at birth of their birth cohort. This measure of health relies on a normative criterion according to which the longer you live, the better is your well-being. This proxy of father/mother's health status is not interpretable in health or in demographics terms but can be compared to life span or longevity, which is highly correlated with a good health status. In this context, we consider that accidental deaths are insignificant.

From a technical point of view, this health indicator of relative longevity differs with gender. Its construction relies on parents' age at death and year of birth. The latter information is not available in the dataset and is thus estimated from the descendant's age and information on maternity and paternity in the 20^{th} century from another source (Daguet, 2002). This estimation is refined by taking into account the descendant's birth position, namely whether the descendant is the eldest or not. Thus, for respondents who were not the eldest of their siblings, their mother's year of birth and respectively their father's year of birth are estimated by subtracting from respondent's year of birth, the average age at maternity and respectively paternity for the same year. Likewise, for the eldest, the mother's year of birth is estimated in the same way but uses the average age at

[13]For a review of literature on this subject, we refer to Cournil and Kirkwood (2001) and the twenty studies they quote.

delivery for first births in the same year. As for the father's year of birth, it is estimated in this case by applying the difference between average age at maternity and average age at paternity to the average age for the first baby birth. Finally, we estimate parents' longevity as the difference between actual longevity and life expectancy at birth (Vallin & Meslé, 2001). The relative longevity of both parents is described in the figure 5.5 and equals on average 14.97 years for mothers and 22.33 years for fathers.

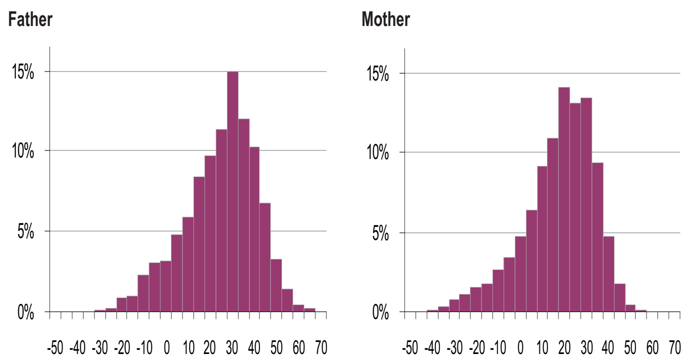

Figure 5.5: Distributions of the relative longevity for fathers and mothers

Conversely to life expectancy which equals the average age at death of a whole generation, this indicator of relative longevity concerns a selected population of men and women, which has at least survived until giving birth to a first baby.[14]

5.4 A first approach in terms of stochastic dominance

The first approach is an analysis of equality of opportunities in health and consists of a comparison of distribution of respondents' health status according to their family and social background. This approach does not limit the definition of equality of opportunity to a simple equality of the average health status which is conditional on family and social background. Indeed, it also studies the effect of family and social background on the whole distribution of health. From the 5-points health status variable, distributions of health

[14]In order to validate our process of estimation for parents' year of birth, we have compared the estimated year of birth with the actual year of birth for parents who were still alive in the original sample. The mean average difference between these two elements equals then three years for fathers and one year for mothers. For fathers, this bias is not correlated to their social status. However, the multiple average comparisons show a significant difference for mothers who are farmers: on average their estimated year of birth is one year later than their actual year of birth. However, couples of farmers are known to have on average more children (Mazuy, 2002; Toulemon, 2003). Thus, we can assume that the average age for the first baby of farmers is higher than the one of other mothers. This bias thus leads to an underestimation of relative longevity for mothers who were farmers. Nevertheless, our results do not provide evidence of any specific effect of this social category.

status are constructed conditional on family and social background, using dominance tests based on a conjunction of *Kolmogorov-Smirnov* unilateral tests.

5.4.1 Dominance according to parents' relative longevity

The hypothesis of inequality of opportunity in health according to parents' health is tested by building distributions of health status conditional on each parent's relative longevity. In order to rely on comprehensive numbers of observations, longevity is considered as a binary variable opposing parents having a relative longevity in the first quartile, i.e. those in poor health to others parents.

This analysis does not conclude that there are inequalities of opportunity in health according to either parents' health existing. Indeed, there is no stochastic dominance between distributions of parents' health as the cumulative distribution functions illustrated in the figure 5.6 clearly overlap.

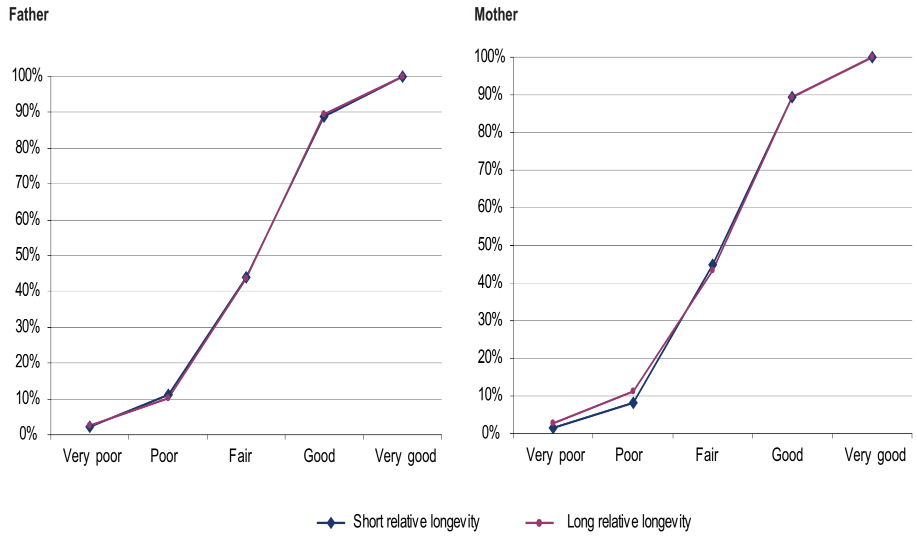

Figure 5.6: Distributions of the respondent's self-assessed health in adulthood according to parents' longevity

Considering that health status changes with age particularly among older sample, we include the effect of each respondent's age on their health in adulthood and do the same analysis separately on three age classes: the 49-60 years old, 61-68 years old and 69 years old and more. When considering the father's health, similar results are found. Nevertheless, the hypothesis of transmission of health is confirmed for descendants aged 61-68 years old when considering the mother's health. The distribution of self-assessed health of individuals of this age who were born to a mother prematurely deceased is therefore dominated at first-order by the distribution of self-assessed health of other individuals.

5.4.2 Dominance according to social background

We then observe the distribution of health status according to social background.

Inequalities of opportunity according to the father's socioeconomic status

The figure 5.7 represents the cumulative distribution function of descendants' self-assessed health conditional on their father's social status. It emphasises that respondents born to "senior managers and professionals" or "technicians and associate professionals" and "armed forces" are more likely to report a good health status than a respondent born to "skilled agricultural and fishery workers", "craftsmen and skilled workers" or "elementary occupations and unskilled workers". Indeed, the cumulative proportion of individuals from favoured social background in very poor and poor health is smaller than the cumulative proportion of those born in socially disadvantaged families. Therefore, the respondent's health is better when his father had a higher socioeconomic position.

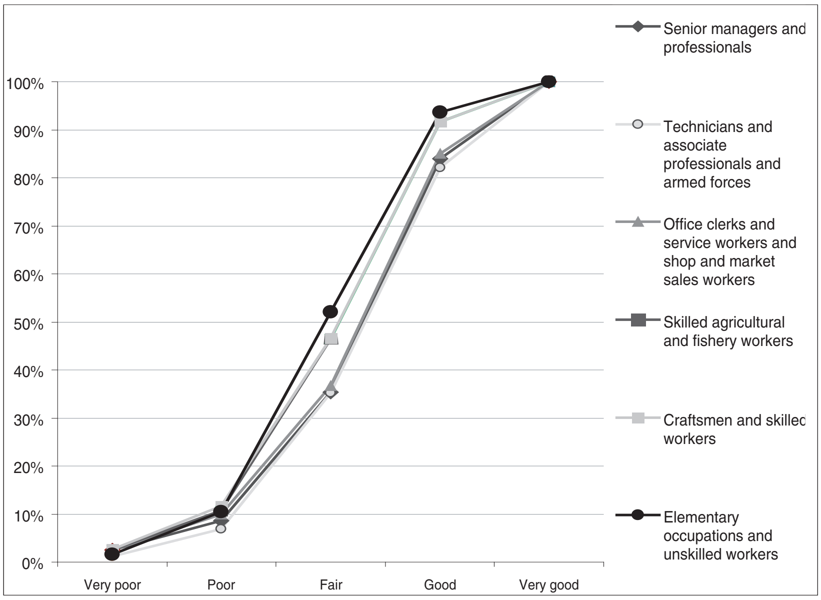

Figure 5.7: Distribution of the respondent's self-assessed health in adulthood according to the father's socioeconomic status

The *Kolmogorov-Smirnov* unilateral tests in the table 5.1[15] confirm the existence of inequalities of opportunity in health according to the father's social status. The distribution of health of respondents whose father belonged to "senior managers and professionals" or "technicians and associate professionals" and "armed forces" significantly dominate the distributions of health of respondents whose father belonged to "skilled agricultural and

[15] Explanation of the table: the result of the unilateral *Kolmogorov-Smirnov* test is read in row. The health distribution of a descendant whose father is in "senior managers and professionals" significantly dominates the one of "skilled agricultural and fishery workers" as p-value=0.0189.

fishery workers", "craftsmen and skilled workers" or "elementary occupations and unskilled workers".

	Senior managers and professionals	Technicians associate prof. armed forces	Clerks service sales workers	Skilled agricultural fishery workers	Craftsmen skilled workers	Elementary occupations unsk. workers
Senior managers professionals		1	0,9708	0,0189**	0,0123**	0,0105**
Technicians associate prof. armed forces	0,9318		0,8755	0,0407**	0,0308**	0,0171**
Clerks service sales workers	1	1		0,1576	0,1382	0.0562*
Skilled agricultural fishery workers	1	1	1		0,9535	0,5453
Craftsmen skilled workers	1	1	1	1		0,5388
Elementary occupations unsk. workers	0,9876	1	0,9901	0,9844	0,9643	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 5.1: P-value of *Kolmogorov-Smirnov* test related to the father's socioeconomic status

Moreover, the results show that the distribution of health in adulthood of respondents born to office clerks or service workers dominates significantly the distribution of health of those born to unskilled workers.

Inequalities of opportunity according to the mother's socioeconomic status

The results are similar for the mother's socioeconomic status (cf. figure 5.8 and table 5.2).

	Senior managers prof. Technicians	Clerks service sales workers	Skilled agricultural fishery workers	Craft and related workers	Elementary occupations unsk. workers	Homemakers
Senior managers professionals Technicians		0,7942	0,0762*	0,0444**	0,0051***	0,0604*
Clerks service sales workers	0,6006		0,0853*	0,0488**	0,0064*	0,072*
Skilled agricultural fishery workers	0,994	1		0,809	0,387	0,992
Craft and rel. workers	0,914	0,993	0,899		0,772	0,942
Elementary occupations unsk. workers	0,985	1	0,984	1		1
Homemakers	0,978	0,994	0,601	0,635	0,184	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 5.2: P-value of *Kolmogorov-Smirnov* test related to the mother's socioeconomic status

The distribution of health status of individuals born to a mother in the groups "senior managers, professionals and technicians" or "office clerks and service workers" dominates significantly the distribution of health of those born to a mother who belonged to any other social category. Therefore, the descendant's health is better if his mother had a higher socioeconomic position.

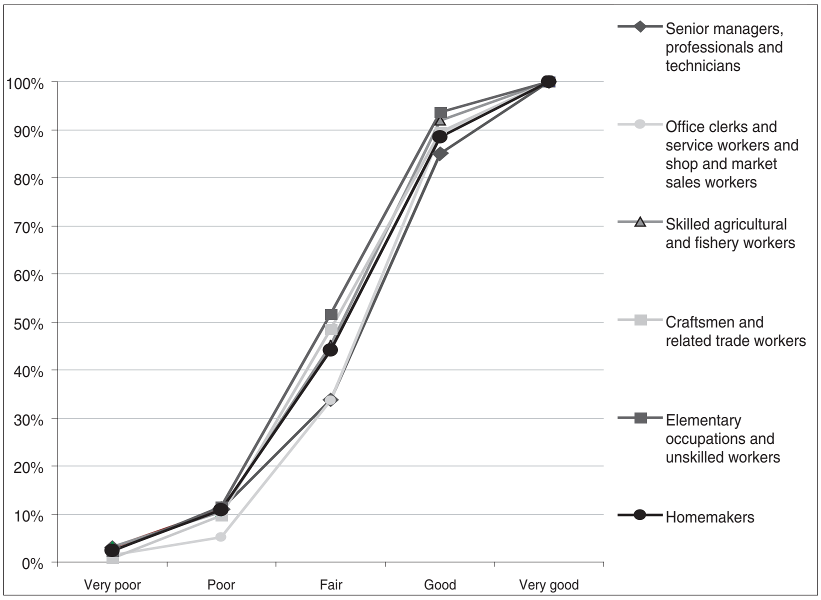

Figure 5.8: Distributions of the respondent's self-assessed health in adulthood according to the mother's socioeconomic status (*2004/05 SHARE*)

This first non-parametric approach emphasises the existence of inequalities of opportunity in health according to social background for individuals from a higher social background and, to a lesser extent to mother's health. These two elements represent circumstances which are independent from individual responsibility.

5.4.3 Dominance according to current socioeconomic status

Traditionally, studies which are interested in social inequalities in health analyse the effects of current social conditions on health in adulthood. In the final section, we consider the descendant's social status as a conditional variable in order to test social inequalities in health and to compare them to inequalities of opportunity in health as previously shown.

This analysis confirms the existence of social inequalities in health according to current social status. Indeed, the distribution of self-assessed health of "senior managers and professionals" and "technicians and associate professionals and armed forces" dominates the distribution of self-assessed health of "office clerks service workers", "skilled agricultural and fishery workers", "craftsmen and skilled workers" and "elementary occupations and unskilled workers".

Similarly, "office clerks and service and shop workers" have a higher probability of being in very good health than "skilled agricultural and fishery workers", "craftsmen and skilled workers" and "elementary occupations and unskilled workers" (cf. figure 5.9 and table 5.3).

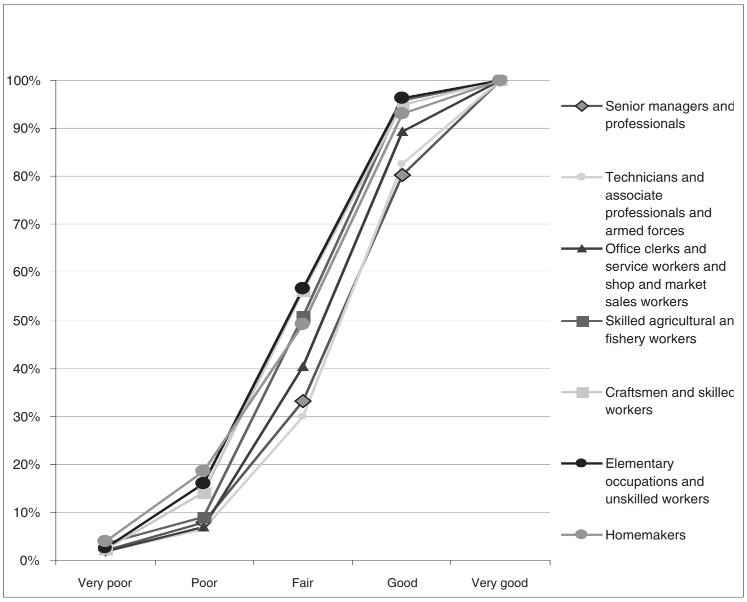

Figure 5.9: Distributions of the respondent's self-assessed health in adulthood according to the respondent's socioeconomic status (*2004/05 SHARE*)

	Senior managers profession	Technicians associate prof. armed forces	Clerks service sales workers	Skilled agricult. fishery workers	Craftmen skilled workers	Elementary occupations unsk. workers	Homemakers
Senior managers professionals		0,858	0,069*	0,003***	<0,0001***	<0,0001***	0,0105**
Technicians associate prof. armed forces	0,724		0,024**	0,0002***	<0,0001***	<0,0001***	0,001***
Clerks service sales workers	0,975	0,999		0,106	0,0001***	0,001***	0,077*
Skilled agricultural fishery workers	1	1	1		0,564	0,465	0,309
Craft and rel. workers	1	1	1	0,973	x	0,922	0,418
Elementary occupations unsk. workers	1	1	1	0,986	1		0,909
Homemakers	1	1	1	0,906	0,706	0,419	

Significance levels: * (10%), ** (5%) and *** (1%)

Table 5.3: P-value of *Kolmogorov-Smirnov* test related to the respondent's socioeconomic status

These social inequalities in health are thus significant and are more pronounced than inequalities of opportunity in health when related to the father's social status. Indeed, the number of significant unilateral tests is higher for distributions of health status in adulthood when it is conditional on current social status rather than on the father's social status. Nevertheless, inequalities of opportunity due to family and social background are not negligible and deserve to be considered. A parametric approach would confirm these first results.

5.5 A second approach using regression analyses

The second approach aims to explain respondents' health status according to family background and relies on ordered Logit regression estimations, which appear to be worthwhile for this approach[16].

We shall consider h_i, the self-assessed health of the descendant i and h_i^* the latent variable which represents "true" health according to which the descendant i self-assesses his health. The variable h_i is a discrete dependent variable that takes multinomial ordered values from 1 corresponding to very good self-assessed health to 5, very poor self-assessed health. We consider various regression models to test the probability of declaring a poor health status. These models gradually add up the descendant's social background, his parents' relative longevity and his own social status. Furthermore, age in five year groups and gender are introduced as control variables.

5.5.1 Influence of social background and parents' relative longevity

The first model estimates the impact of parents' social status and relative longevity on a decreasing health status. We shall denote parents' social status as SES_{fath} and SES_{moth} and their respective relative longevity as $Long_{fath}$ and $Long_{moth}$. The model 1 is written as follows.

$$\begin{aligned} h_i^* = \beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_4 \times SES_{moth} \\ + \beta_5 \times Long_{fath} + \beta_6 \times Long_{moth} + u \end{aligned} \tag{5.1}$$

The results show that the probability of self-assessing a poor health in adulthood decreases with parents' socioeconomic status (cf. Table 5.4, model 1). An individual born to a father whose occupation is either "senior managers and professionals" or "technicians and associate professionals" and "armed forces" or "office clerks and service workers", has a significantly lower probability of poor health status than those whose father have an

[16]We have used the Brant test of the parallel regression assumption (Brant, 1990) to know whether the assumption of parallel slopes of ordered logit models is confirmed. This test relies on Wald Tests to test the hypothesis that the coefficients in each independent variable are constant across categories of the dependent variable. This test is useful in two respects. Firstly, it indicates that we should perhaps estimate a generalised logit model, and secondly, it suggests what variables may be used in determining the thresholds. The Brant test and results concerning our sample are presented in appendix B.

elementary occupation, after controlling for age and gender. These results match exactly with those emphasised by the dominance approach. Considering mothers' socioeconomic status, a respondent whose mother had an elementary occupation always has a higher probability of poor health status in adulthood than someone born to a mother who was a homemaker. Moreover, individuals whose parents had a higher longevity than other people in their generation have a significantly reduced risk of poor health.

Considering the construction of parents' relative longevity, there could be a correlation with descendant's age implying a biased estimation. In order to validate our construction of parents' health status, we test an alternative to model 1 involving parents' age at death, instead of their relative longevity. Age includes the generation effects, which then influence the life expectancy:

$$h_i^* = \beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_4 \times SES_{moth} + \beta_5^{'} \times Age_{fath}^{death} + \beta_6^{'} \times Age_{moth}^{death} \quad (5.2)$$

The odds ratio in this alternative model are very close to those in model 1 (cf. Table 5.4, model 1bis). It confirms that relative longevity is a valid instrument and does not introduce any bias in our analysis.

5.5.2 Influence of social background, parents' relative longevity and current socioeconomic status

As the specification of model 1 is limited to few variables, this model can be called into question. Indeed, some omitted variables such as the descendant's education level or his social status can induce an endogeneity bias of parents' social status on the self-assessed health in adulthood.

We shall now denote the descendant's education as $Educ_i$, such as model 2 is:

$$h_i^* = \beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_4 \times SES_{moth} + \beta_5 \times Long_{fath} + \beta_6 \times Long_{moth} + \beta_7 \times Educ_i \quad (5.3)$$

This model tests whether the influence of social background shown in the first model comes from its direct effect on health in adulthood or from its indirect effect through the descendant's education (cf. Table 5.4, model 2).

We observe *ceteris paribus* that education significantly influences health status: the higher the education level, the lower the risk of poor health. In addition, the introduction of education level modifies previous results: the effect of the father's socioeconomic status on descendant's health is removed; it is indirect and comes from the respondent's education level, which reminds the *pathway model*.

On the contrary, the influence of the mother's socioeconomic status on health status in adulthood persists: the probability of being in poor health in adulthood is higher for individuals born to mothers in elementary occupations than homemakers. This direct impact confirms the *latency hypothesis*. This latter effect can be interpreted as either the influence of current living standards or the influence of the mother's academic standard on health in adulthood. Furthermore, the introduction of education reduces the significance of odds ratios related to parents' health. This result suggests that education could reduce the influence of parents' health, i.e. the transmission of intergenerational inequalities in health. A higher education level would thus be able to protect health, because of a lower reproduction of poor family habits or an improved awareness of health transmitted problem such as genetic screening. A third model is estimated and adds descendant's social status, SES_i:

$$h_i^* = \beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_4 \times SES_{moth}$$
$$+\beta_5 \times Long_{fath} + \beta_6 \times Long_{moth} + \beta_7 \times Educ_i + \beta_8 \times SES_i \quad (5.4)$$

This added variable introduces the need to correct for endogeneity due to social background of respondents in the model. Despite the introduction of SES_i, social background nevertheless has a direct effect on health in adulthood (cf. Table 5.4, model 3). The probability of assessing a poor health is higher when respondents are unskilled workers.

In regard to parents' health, we conclude that an increase of one year of this variable for both parents, decreases, *ceteris paribus*, the probability of being in very poor health in adulthood of 0.52% for the father (respectively 0.55% for the mother); the probability of being in very poor or poor health in adulthood of 0.47% for the father (resp. 0.49% for the mother); the probability of being in very poor, poor or fair health in adulthood of 0.25% for the father (resp. 0.26% for the mother); the probability of being in very, poor, fair or good health in adulthood of 0.03% for both parents. For a reference individual[17], having a mother in elementary occupations instead of homemaker increases the probability of being in very poor health to 35%; the probability of being in poor or very poor health to 31%; the probability of being in fair, poor or very poor health to 15% and the probability of being in good, fair, poor or very poor health to 1.6%.

[17] The relative longevity of a reference individual's parents equals average values, i.e 22 years old for the father and 15 for the mother

		Poor self-assessed Health in 5 categories			
Variables	Freq.	Model 1	Model 1bis†	Model 2	Model 3
Gender					
Woman	999	0.811**	0.81**	0.747***	0.75***
Man	784	ref.			
Age					
49-54 years old	162	1.206	1.144	1.156	1.16
55-59 years old	235	ref.			
60-64 years old	246	1.502**	1.517**	1.443**	1.495**
65-69 years old	302	1.679***	1.711***	1.686***	1.756***
70-74 years old	305	2.399***	2.395	2.237***	2.278***
75-79 years old	257	3.51***	3.584	3.316***	3.506***
80-84 years old	192	4.507***	4.51	4.223***	4.554***
>=85 years old	84	7.029***	6.771	6.427***	7.211***
Father's occupation					
Senior managers and professionals	243	0.568**	0.601**	0.829	0.906
Technicians and associate professionals and armed forces	173	0.509***	0.529***	0.757	0.863
Office clerks and service workers and shop and market sales workers	120	0.583**	0.614**	0.808	0.84
Skilled agricultural and fishery workers	482	0.805	0.839	0.807	0.813
Craftsmen and skilled workers	640	0.882	0.879	0.951	0.98
Elementary occupations and Unskilled workers	125	ref.			
Mother's occupation					
Senior managers, professionals and technicians	154	0.869	0.857	1.031	1.095
Office clerks and service workers and shop and market sales workers	137	0.794	0.769	0.836	0.888
Skilled agricultural and fishery workers	286	1.014	1.046	0.985	0.965
Craft and related trades workers	134	1.019	0.99	1.039	1.03
Elementary occupations and unskilled workers	174	1.331*	1.304*	1.335*	1.361**
Homemakers	898	ref.			
Father's health					
Father's relative longevity	1783	0.991***	0.991***	0.994**	0.994**
Mother's health					
Mother's relative longevity	1783	0.992***	0.992***	0.995*	0.994*
Education level					
Elementary level diploma	555			0.624***	0.673***
Secondary level diplomas	481			0.5***	0.639***
Baccalauréat (A-levels)	341			0.271***	0.412***
No diploma	406	ref.			
Descendant's occupation					
Senior managers and professionals	280				0.456***
Technicians and associate professionals and armed forces	311				0.396***
Office clerks and service workers and shop and market sales workers	387				0.595***
Skilled agricultural and fishery workers	142				0.73
Craftsmen and skilled workers	339				0.856
Elementary occupations and unskilled workers	194	ref.			
People taking care of home and family	130				0.747
Model quality					
Score Test for the Proportional Odds Assumption (P-value)		0.2096	0.203	0.1991	0.1230
AIC		4318.585	4317.683	4259.021	4238.319
Concordant pairs percent		64.8	64.9	67.0	68.2
Adjusted R2		0.1131	0.1136	0.1483	0.1654

Significance levels: * (10%), ** (5%) and *** (1%) † with parents age at death

Table 5.4: Ordered Logit regression models: odds ratio of poor health (*2004/05 SHARE*).

5.5.3 Endogeneity test of the social status in adulthood

The third model does not permit removing the potential endogeneity bias due to the descendant's socioeconomic status, namely SES_i. Indeed, there could exist variables influencing both health and social statuses in adulthood. For instance, the health status at 20 years old could have influenced the occupational status of an individual, but this information is not available in our data. To test this potential endogeneity, we use Lollivier's method (2001).

This method relies on the estimation of a bivariate probit model in two equations: the first one explaining health status in adulthood and a second one explaining the probability of having a higher social status. An instrument variable of the descendant's social status is thus required in order to identify the model. Among the variables of our dataset, height in adulthood is the only instrument variable: it is positively correlated to social status and insignificantly linked to health[18]. Even if some references in the literature show significant correlation between height and self-assessed health (Silventoinen *et al.*, 2007), there are other references emphasising the relationship between height and professional career (Herpin, 2003). Height is thus both a statistical and an intuitive instrument of the respondent's socioeconomic status.

We shall denote H_i, the descendant self-assessed health considered as dichotomous[19] and SES_{sup_i}, a dichotomous variable for descendant's health status[20], the last model is threfore composed of two equations, the first equation tests the probability of declaring a poor health status H_i whereas the second equation tests the probability of having a higher social status.

$$P(H_i = 1/X) = F\,(\beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_4 \times SES_{moth} + \beta_5 \times Long_{fath} + \beta_6 \times Long_{moth} + \beta_7 \times Educ_i + \beta_8 \times SES^*_{sup_i} + u) \quad (5.5)$$

$$P(SES_{sup_i} = 1/X) = F\,(\beta_1 \times Sex + \beta_2 \times Age + \beta_3 \times SES_{fath} + \beta_5 \times Long_{fath} + \beta_6 \times Long_{moth} + \beta_7 \times Educ_i + \beta_8 \times Height_i + e) \quad (5.6)$$

Where F is a cumulative distribution function associated to the probit distribution. Adult height (in centimeters) is found significantly correlated to the probability of having a higher social status (cf. Table 5.5).

The second equation emphasises the link between social conditions in adulthood and social background: the parents' socioeconomic status significantly increases the probability of reaching a higher social position, *ceteris paribus*.

[18]The correlation between "having a higher social status" and "being taller than average" significantly equals 0.17 for men and 0,05 for women

[19]very good, good versus fair, poor and very poor health

[20]"Senior managers and professionals", "technicians and associate professionals" and "armed forces" compose the higher social group versus all the other lower social statuses.

Variables	Probability to have a poor health status	Probability to have a higher social status
Gender		
Woman	-0,213**	-0,292***
Man	ref.	
Age		
49-54 years old	0,091	0,043
55-59 years old	ref.	
60-64 years old	0,26*	0,435***
65-69 years old	0,35***	0,407***
70-74 years old	0,57***	0,25*
75-79 years old	0,807***	0,338**
80-84 years old	0,956***	0,522***
>=85 years old	1,18***	0,8***
Father's occupation		
Senior managers and professionals, technicians and associate professionals and armed forces	0,043	0,461***
Others	ref.	
Mother's occupation		
Senior managers and professionals, technicians and associate professionals	-0,025	0,267**
Others	ref.	
Father's health		
Father's relative longevity	-0,006***	-0,001
Mother's health		
Mother's relative longevity	-0,004**	-0,001
Descendant's occupation		
Senior managers and professionals, technicians and associate professionals and armed forces	-0,476	
Others	ref.	
Education level		
Elementary level diploma	-0,292***	0,363***
Secondary level diploma	-0,315	0,98***
Baccalauréat (A-levels)	-0,443	2,1***
No diploma	ref.	
Height		
Descendant's height		0,015***
Intercept	-0,125	-4,04***
Rho	-0,097	0,371
Model quality		
Separated model log-likelihood	-1109.6328	-794.28352
Simultaneous model log- likelihood	-1903.8824	
Endogeneity test rho=0	chi2(1) = 0.06786	Prob > chi2 = 0.7945

Significance levels: * (10%), ** (5%) and *** (1%)

Table 5.5: Bi-probit estimation, Lollivier test (*2004/05 SHARE*)

The Lollivier endogeneity test consists of comparing log-likelihood values from the two previous probit models estimated simultaneously and separately. The separated estimation implies the following log-likelihood:

for equation (5.5), $L_1 = -1109.057$

for equation (5.6), $L_2 = -794.106$

for the simultaneous model, $L = -1903.102$.

The endogeneity statistic test equals $(-2L_1) + (-2L_2) - (-2L) = 0.123$. It follows a $Chi - square(1)$ and therefore it does not reject the null hypothesis in our context (0.123<3.84). The two error terms e and u are not correlated, which implies that a simultaneous estimation is useless.

The descendant's social status appears exogenous when explaining health in adulthood according to social background and parents' health. Equation (5.4) thus offers a sufficiently explained specification as there are no unobserved variable, which implies a correlation between descendant's social status and the error term.

5.6 A third approach using concentration indices

Most of the recent empirical works on inequalities in health use the concentration index to evaluate inequalities in health. Generally, the concentration curve related to this index graphs the cumulative proportion of health against the cumulative proportion of the population ranked by income. Nevertheless, the original role of the concentration index is to describe the relationship existing between two distributions regardless of the variables. In this section, we propose to compute four concentration indices measuring inequalities in health in adulthood related to the relative longevity of each parent and related to their respective socioeconomic status. The calculation of concentration indices requires a measurement of respondent's health, which is valid for the analysis as seen in previous chapters. In this context, we follow the method used in chapter 4 and cardinalise the 5-categories self-assessed health with the distribution of SF6D corresponding to the French SF36. Whereas the calculation of the health concentration index related to parents' health is straightforward as the two distributions involved in the concentration index are continuous, the second application relies on an assumption on the ranking of parents' socioeconomic status.

5.6.1 Measurement of health: cardinalisation of self-assessed health with SF6D

In the same way as it has been done in chapter 4, the measurement of health relies on the SF6D utility algorithm applied to the French SF36 questionnaire from the 2003 National Health Survey. We assume a stable mapping from SF6D to the reported self-assessed health in the French part of SHARE. Therefore, we compute the cumulative frequency of observations for each category of self-assessed health in SHARE and find the corresponding quantiles of SF6D. The table 5.6 presents the matching thresholds.

Self-assessed health	Cum. frequency	SF6D quantiles
Very poor	42	0,501
Poor	185	0,566
Fair	778	0,647
Good	1594	0,772
Very good	1783	0,948

Table 5.6: Cumulative frequencies of self-assessed health and quantiles of SF6D

The resulting measurement of health allows us to compute inequality measures using interval regression models.

5.6.2 Inequalities in health related to parents' relative longevity

We measure inequalities in health in adulthood over the distribution of the father's (respectively mother's) relative longevity. The concentration index requires a ranking variable for the population. We use the longevity of each parent and rank individuals from the most disadvantaged in father's (respectively mother's) relative longevity to the least disadvantaged. We compute two concentration indices of inequality in health such as the two related concentration curves plot on the X-axis, the cumulated proportion of individuals according to father's longevity (respectively mother's) with on the Y-axis the SF6D rescaled self-assessed health of the respondent.

These concentration indices rely on the estimation of two linear regression models explaining self-assessed health. The first model involves father's longevity whereas the second model concerns mother's longevity. This can be written as follows.

$$h_i = \beta_1 Sex + \beta_2 Age + \beta_3 Long_{fath} + \beta_4 SES_i \tag{5.7}$$

$$h_i = \beta_1 Sex + \beta_2 Age + \beta_3 Long_{moth} + \beta_4 SES_i \tag{5.8}$$

We assume that health status in adulthood h_i of individual i is defined according to various regressors, which are demographics, education level and social status. We carry out two different interval regression models. The table B.4 in appendix B presents results of these two interval regressions. It is noteworthy that results are similar to those in section 5.5 qualitatively and according to the degree of significance. Consequently, commenting these results is of limited interest. Nevertheless, these results permit giving an overall picture of results on global inequality, which is presented in the summary table 5.7. The inequality in health in adulthood over parents' longevity, called CI is presented as the sum of two elements: the inequality driven by policy irrelevant characteristics, called CI^*, and the inequality from policy relevant characteristics, called I. This latter inequality I measures the inequality of opportunity in health, which is also the inequity in health, it can also be composed in two elements: inequality related to individual social characteristics and inequality in the distribution of parents' longevity.

The two global concentration indices describe inequalities in health status favouring respondents born to parents with a lower longevity as shown by positive CIs. Nevertheless, it is important to emphasise that these are raw concentration indices, which do not standardise on demographics. In our sample, demographics are playing a relevant role as shown by both significant odds ratios in table 5.4 and their high contributions to CI. The standardisation on demographics is thus useful to stress the *pure* value of inequalities of opportunity in health. Indeed, the effects of age on health status are strong when people get older and it is important to consider a concentration index which controls for age of respondent.

The standardised inequality, namely I displays an inequality of opportunity in health favouring respondents born to a father, respectively a mother with a higher longevity. The inequality of opportunity related to father's health is slightly higher than the inequality related to mother's health. In the previous parametric approach, the odds ratio related to the father's longevity in the model 3 (table 5.4) is also slightly higher than odds ratio related to the mother's health. As the concentration indices are defined on a ranking by relative longevity, it is expected to observe a higher contribution for relative longevity than for social characteristics to the global concentration indices. These results confirm our previous conclusions. Due to the very original ranking variable used, it is difficult to compare these values with concentration indices computed in chapter 4. Indeed, it is more widespread to rank individuals on their social characteristics as it is done in the following section.

Variables	Father's longevity	Mother's longevity
CI global	-0,0017	-0,0032
Policy irrelevant characteristics		
Demographics CI^*	-0,0068	-0,0073
Policy relevant characteristics		
$I = CI - CI^*$	0,0051	0,0041
Relative Longevity	0,0047	0,0036
Social characteristics	0,0004	0,0006

Table 5.7: Decomposition of concentration indices of inequalities in health related to parents' relative longevity

5.6.3 A pseudo health concentration index according to parents' socioeconomic status

Concentration index is generally computed on two continuous distributions; however, ranked grouped variables can also be used (Wagstaff *et al.*, 1991). Such a concentration index is analogous to a pseudo Gini index with an ordered variable (Preston *et al.*, 1981), we shall call it a pseudo concentration index. Traditionally, social inequalities in health analyse effects of current social conditions on health. Therefore, in this section we present an original use of these tools. At first, we compute a pseudo health concentration index of inequality in health according to respondent's socioeconomic status. This first application will be compared to concentration indices computed in chapter 4. In a second step, we will compute pseudo-concentration indices according to parents' socioeconomic status. Whereas the pseudo-concentration index related to individual's social status represents social inequalities in health, pseudo-concentration indices related to parents' social status give a measure of inequalities of opportunity in health.

Socioeconomic status: a ranking variable?

Our previous results on stochastic dominance help to define a consistent ranking among both respondent's and his parents' social classes.

From figure 5.9 and table 5.3, we can define a consistent ranking of socioeconomic statuses for respondents. "Senior managers" are the most socially advantaged and "elementary occupations and unskilled workers" are the least socially advantaged. We position "homemakers" as the second least advantaged, then ranking follows ISCO classification.

As for fathers, table 5.1 and figure 5.7 describe clearly the three highest social classes: 1. "senior managers and professionals"; 2. "technicians and associate professionals, armed forces" ; 3. "office clerks, service workers and shop and market sales workers", and the lowest social class: 6. "elementary occupations and unskilled workers". Nevertheless, it is difficult to decide which is socially higher, "skilled agricultural and fishery workers" or "craftsmen and skilled workers". That is the reason why we follow the ISCO classification which ranks craftsmen after agricultural workers.

As for mothers, the ranking of social classes is clearly shown by table 5.2 and figure 5.8 and is as follows: 1. "senior managers, professionals and technicians"; 2. "office clerks and service workers and shop and market sales workers"; 3. "homemakers"; 4. "skilled agricultural and fishery workers"; 5. "craft and related trades workers" and 6. "elementary occupations and unskilled workers".

In this context, the construction of three pseudo-concentration indices relies on the construction of concentration curves that plot on the X-axis, the cumulated proportion of individuals ranked from the lowest social classes with on the Y-axis the SF6D rescaled self-assessed health of the respondent.

Inequalities in health over socioeconomic status

The computation of concentration indices rely on three interval regression models explaining self-assessed health in adulthood mapped on SF6D. Regression results are presented in table B.5 in appendix B. Relationships of regressors with self-assessed health are qualitatively and significatively similar to those in section 5.5. Therefore we move directly to the description of the global inequality in health over socioeconomic status and social background, which is presented in the summary table 5.8. This global inequality in health in adulthood, namely CI is decomposed in two elements: the inequality driven by demographic characteristics, namely CI^*, and the inequality from relevant social characteristics, namely I. When individuals are ranked according to their own socioeconomic status, the index I measures the inequity in health over social status. When they are ranked according to their father's or mother's socioeconomic status, I measures the inequality of opportunity in health.

Variables	Individual's SES	Father's SES	Mother's SES
CI global	0,0164	0,00761	0,0066
Policy irrelevant characteristics			
Demographics CI^*	0,0003	-0,00003	0,0004
Policy relevant characteristics			
$I = CI - CI^*$	0,0161	0,00763	0,0062
Ordered SES	0,0100	0,00201	0,0019
Social characteristics	0,0062	0,00562	0,0043

Table 5.8: Decomposition of concentration indices of inequalities in health related to social status and social background

The ranking over individual's socioeconomic status displays an inequality in health favouring higher socioeconomic statuses. When self-assessed health in five categories was cardinalised using SF6D in chapter 4, the income-related inequality in health was $CI = 0,0054$, therefore the concentration index is higher over socioeconomic statuses than over income. According to Chenu (2000), socioeconomic classes always display strong differences in health. In the context of a concentration index, this higher value comes from the categorical ranking variable which gives a less precise ranking than the continuous income and emphasises differences between extreme groups. Similarly, inequity in health over social status is higher than inequity in health over income.

The ranking over social background also shows an inequality in health favouring individuals born to socially advantaged families. Nevertheless, it is noteworthy that concentration indices related to parents' social status are lower than the concentration related to individual's social status. This result is in line with the previous stochastic dominance analysis in section 5.4.3. Social inequalities in health were found more pronounced than inequalities of opportunity in health when related to the father's social status. More precisely, for both fathers and mothers, contributions of ordered socioeconomic status to the global inequality are substantially lower than contributions of individual's social characteristics. Indeed, father's social status contributes for $26,5\%$ (respectively $28,6\%$ for the mother) whereas social characteristics contribute for $73,9\%$ (respectively $65,5\%$). Moreover, inequalities in health are higher when the father's social class is considered. These results confirm previous conclusions and emphasise the existence of both inequalities in health and inequalities of opportunity in health in adulthood.

5.7 Conclusion

The evidence from this analysis indicates inequalities of opportunity in health in adulthood according to social background and parents' longevity. Whereas the mother's socioeconomic status has a direct effect on health in adulthood, as described by the *latency hypothesis*, the father's socioeconomic status has an indirect effect through the descendant's social status in accordance with the *pathways hypothesis*. Moreover, the hypothesis of health transmission from one generation to the next is shown as there is a direct effect of each parent's relative longevity on the health of their descendants in adulthood. The final analysis with concentration indices confirms previous results and shows higher inequalities in the distribution of the father's longevity and socioeconomic status. As a consequence, the three ways through which family background can influence health in adulthood, as has been shown, are involved in the explanation of inequalities of opportunity in health in France.

Our results rely on three alternative approaches. The non-parametric approach gives results in terms of stochastic dominance at first-order. These results are very robust because they come from a free-model. The parametric approach confirms and refines results by reasoning *ceteris paribus*. The concentration indices offer a more synthetical point of view on the existing inequalities and propose an atypical use of this standard tool for measurement of inequalities in health. To our knowledge, the use of three different approaches to analyse the same research question has never been done. It is particularly relevant because each measurement tool has its own limitations and the multi-approach analysis allows researchers to reinforce their conclusions.

The use of the self-assessed health to measure the respondent's health could be criticised as this variable can suffer from reporting heterogeneity as reviewed in chapter 1. Nevertheless, our study sample concerns older individuals and they are less likely to mis-report their health status, especially after retirement (Bryant *et al.*, 2000). Moreover, a similar analysis has been conducted using functional limitations in daily life as a definition of health in adulthood. This analysis shows analogous results.

Our analysis presents some limitations, particularly linked to data.

- Due to our restricted sample of variables, we could not have found a variable that well-instruments the respondent's socioeconomic status.

- The health information concerning parents was limited. Data of better quality would allow a more accurate understanding of the causal pathways between childhood conditions. In particular, we are unable to discern whether transmission of health is due to genetic inheritance or copying parental behaviours (health preferences, risky behaviours). This question is important in an analysis of inequalities of opportunity because from an ethical point of view, inequalities due to genes will not be equivalent to inequalities in social background (Lefranc *et al.*, 2004). In particular, public policies of correction in each context would be very different, too. Furthermore, the effect of parents' health could also be explained by a common family characteristic influencing the health status of all the members in the family. For example, a similar exposure to either a risky geographical environment (radioactive, environmental pollution) or a similar sanitary risk or a socially disadvantaged context would suggest similar health statuses within a family generation.

- This analysis has been conducted for one year and the ideal design would be to follow a cohort from birth into adulthood. In particular, data does not allow us to explore the relationship between parents' socioeconomic status and children's health, and afterwards the relationship between childhood and adulthood. Nevertheless, the gradient in health status according to which wealthier people have better health and live longer is observed in adulthood but has antecedents in childhood.

- Moreover, sociological studies have shown that the relationship of socioeconomic status in childhood with both socioeconomic status and health-related outcomes in adulthood differs according to the country (Breen, 2005). For example, the latter association would be stronger in UK. As a result, the European dimension of SHARE is very interesting in order to evaluate differences in equality of health opportunities according to childhood circumstances.

Regarding these limits, we have introduced in the last IRDES-HHIS survey, a questionnaire on childhood[21]. Despite the general limitations of retrospective reports, as distressed people may recall their life history according to their emotions, this questionnaire has promising perspectives[22] and the reliability of retrospective reports of childhood ex-

[21] Cf. appendix B for a presentation of this questionnaire.
[22] The non-response concerning the questionnaire is low.

periences has already been assessed (Dube *et al.*, 2004). Moreover, as this set of questions is introduced in the same dataset, the indicator developed in chapter 2 could be used to measure individual health status for a new application of this study to the 2006 IRDES-HHIS. It would give a more global health measurement and would be less subjective than self-assessed health.

Therefore, results of this chapter offer various extensions for empirical work at national and European levels.

Conclusion générale

Cette thèse a permis d'étudier les inégalités sociales de santé sur données françaises en fournissant un nouveau concept de mesure de la santé et en employant des méthodes de mesures originales. Les résultats sont d'autant plus robustes qu'ils se sont appuyés à la fois sur des comparaisons méthodologiques et sur des analyses de sensibilité. Les apports de la thèse se situent à deux niveaux : la mesure de la santé, qui a fait l'objet de la première partie, et l'analyse des inégalités sociales de santé, réalisée dans la seconde partie.

La première partie a dressé, à travers le chapitre 1, le cadre conceptuel dans lequel s'inscrit la mesure de la santé et a fourni, dans le chapitre 2, un outil de mesure de la santé qui a respecté ce cadre. L'indicateur de santé construit est pragmatique et global. Il rend compte de la multidimensionnalité de la santé et propose une mesure de la santé tout à fait équilibrée entre la santé, dite subjective et la santé que nous avons qualifiée de moins subjective. Cet indicateur se différencie des outils de mesure de la santé jusqu'ici disponibles au niveau national comme international, par le fait qu'il ne relève pas d'un questionnaire spécifique et s'appuie simplement sur des données d'enquêtes. Cet indicateur ne prétend pas à l'universalité, toutefois sa méthode de construction a l'avantage de pouvoir être aisément reproduite avec d'autres variables de contrôle et sur d'autres échantillons, sous réserve qu'ils disposent d'un report de maladies auxquelles l'indice de sévérité puisse être appliqué. La thèse démontre que cet indicateur peut être utilisé pour des comparaisons simples d'état de santé dans différentes populations, pour calculer de nombreuses statistiques et pour comparer des distributions. Il a en outre pu être utilisé dans l'analyse des inégalités sociales de santé dont les résultats seront rappelés ci-après. Il nous semble cependant que ce nouvel indicateur ouvre des perspectives d'utilisation au-delà de celles proposées dans la thèse. Nous pensons, par exemple, à une analyse des consommations de soins selon le besoin de santé qui ferait intervenir plusieurs indicateurs de morbidité. Dans ce cadre, l'indicateur serait une bonne solution pour éviter des phénomènes d'autocorrélation.

La seconde partie a fourni dans le chapitre 3, un cadre pertinent pour l'analyse des inégalités sur lequel se sont appuyées les analyses empiriques menées aux chapitres 4 et 5.

Le chapitre 3 s'est attaché à redéfinir les critères de l'inégalité unidimensionnelle puis multidimensionnelle. La santé et le revenu représentent deux dimensions du bien-être individuel qu'il est pertinent de considérer conjointement. Si les outils de dominance stochastique permettent de conclure sans ambiguité et de manière robuste à l'existence de différences significatives entre des distributions, les indices de mesure des inégalités viennent compléter ces faits en mesurant l'amplitude de ces différences. Ce chapitre nous a permis de sélectionner la dominance stochastique à l'ordre 1 et l'indice de concentration pour les analyses empiriques qui ont suivi. L'emploi de la dominance stochastique à l'ordre 1, dans le chapitre 5, est justifié par le caractère discret de la variable d'intérêt. Quant à l'intérêt d'utiliser l'indice de concentration aux chapitres 4 et 5, il a relevé à la fois de son aspect synthétique du degré d'inégalité, de sa capacité à être décomposable et de sa comparabilité avec les travaux européens antérieurs.

Le chapitre 4 a élargi les connaissances que nous avions jusqu'à présent des inégalités sociales de santé en France. Nous avons mis en oeuvre une méthode originale, reconnue internationalement, pour évaluer l'ampleur et pour expliquer assez finement les disparités dans la distribution de l'état de santé général. L'utilisation de cette méthode de décomposition de l'indice de concentration est pertinente, car elle tient compte à la fois des relations causales entre les différents facteurs explicatifs introduits dans l'analyse et de leurs liens avec la santé. De cette manière, les résultats ont mis en évidence que non seulement l'inégale distribution du revenu dans la population, mais aussi la forte élasticité de la santé avec le revenu, font de ce paramètre le principal déterminant des inégalités sociales de santé en France en 2004. En outre, nous avons montré que les inégalités s'expliquent dans une moindre mesure par le niveau d'éducation et la catégorie socioprofessionnelle.

Alors que nos résultats montrent des inégalités sociales de santé significatives au détriment des plus pauvres en 2004, ils mettent aussi en évidence une diminution des inégalités sur la période 1998-2004. Cette diminution s'explique notamment par une diminution de la force de l'association existant entre la santé et le revenu et une plus faible inégalité dans la répartition du revenu au sein des groupes sociaux. En outre, la situation sociale des plus âgés de l'échantillon (56-65 ans) se serait améliorée. Ce résultat est sans précédent au regard des nombreuses études établissant une augmentation des inégalités sociales de mortalité ou encore, des inégalités sociales dans la prévalence de certaines maladies en France (de Koninck & Fassin, 2004). Il serait intéressant de répliquer l'analyse sur des données longitudinales disposant d'échantillons plus importants et de variables de revenu plus détaillées, afin de valider ces résultats. L'analyse demande également à être reproduite sur des années ultérieures afin de confirmer la tendance observée.

Dans ce chapitre, nous avons démontré également que le choix d'une mesure de santé n'est pas anodin sur la mesure des inégalités sociales de santé sous-jacente. En effet, l'ampleur de l'inégalité dépend à la fois du nombre de catégories de la variable à cardinaliser et de la distribution de santé choisie pour réaliser cette cardinalisation. Etant donné qu'il n'existe pour le moment aucun consensus en la matière, il nous semble que l'approche consistant à faire appel à plusieurs indicateurs de santé présente l'avantage d'offrir des résultats concordants.

Le chapitre 5 a apporté de nouveaux éléments à l'explication des inégalités sociales de santé à l'âge adulte. En particulier, il a mis en cause les conditions sociales dans l'enfance, approchées par la profession de chacun des parents ainsi que leur longévité. Nous nous sommes appuyés sur trois approches distinctes. Ces approches ont conduit à des résultats similaires, tout en apportant chacune des particularités méthodologiques, ce qui donne du corps à nos conclusions. L'approche non paramétrique a permis de classer, sans ambiguïté, les distributions de santé à l'âge adulte selon les professions du père puis de la mère. Elle n'a cependant pas permis de conclure sur les distributions de santé selon la longévité des parents lorsque celle-ci est considérée en variable discrète. L'approche par un modèle explicatif de l'état de santé à l'âge adulte a permis de distinguer les effets de différentes caractéristiques individuelles et de contrôler les effets observés avec les variables démographiques. Elle a, notamment, mis en évidence l'influence significative de la longévité du père et de la mère sur la santé à l'âge adulte. Enfin, l'approche par indices de concentration a permis d'évaluer l'ampleur des inégalités, de manière synthétique, selon la longévité et selon la catégorie sociale ordonnée. Puis, la méthode de décomposition s'est avérée très fructueuse pour évaluer la contribution à l'inégalité de groupes de variables, comme les caractéristiques socioéconomiques à l'âge adulte ou celles correspondant aux conditions dans l'enfance.

De tels résultats n'ont jamais été montrés concernant la population générale en France. Nous nous sommes appuyés sur le concept d'inégalités des chances pour qualifier ces inégalités dues à des caractéristiques indépendantes de la responsabilité individuelle. En outre, à chaque approche, nous avons fait correspondre une analyse comparée des inégalités sociales de santé, qui relèvent uniquement des caractéristiques individuelles à l'âge adulte. Nous avons alors montré que les inégalités des chances en santé sont moins marquées que celles liées aux caractéristiques actuelles.

La qualité des données de l'Enquête Santé et Protection Sociale sur la morbidité individuelle a permis la construction d'un indicateur de santé cardinal et innovant dans le chapitre 2, qui faisait défaut pour les études sur la santé en France.

Cependant, sous d'autres aspects, les données utilisées dans cette thèse présentent des faiblesses et n'ont pas permis de mener des analyses aussi fines que nous l'aurions souhaité.

Les résultats du chapitre 4 souffrent, en effet, des limites de l'Enquête Santé et Protection Sociale à fournir des données détaillées sur le revenu ou à atteindre les ménages les plus précaires. Par exemple, nos données sous-estiment la proportion d'individus bénéficiaires de la couverture maladie universelle. De même, il a été préférable de limiter l'analyse à la population en âge de travailler.

Dans le chapitre 5, l'analyse a dû s'adapter à l'absence de certaines données comme l'année de naissance des parents, les conditions de vie durant l'enfance ou encore les comportements à risque des individus dans les données de l'enquête SHARE. Nous avons, par exemple, estimé l'année de naissance de chacun des parents à partir de l'année de naissance de l'enquêté et des informations connues sur les âges moyens à la maternité et à la paternité au cours du vingtième siècle. Cependant, nous n'avons pu qu'émettre des hypothèses sur les caractéristiques de santé transmises d'une génération à une autre. En particulier, il aurait été pertinent de pouvoir estimer un modèle expliquant la santé à l'âge adulte de manière plus précise. Nous pensons par exemple aux antécédents médicaux des parents, aux comportements risqués ou préventifs durant l'enfance et au cours de la vie, aux caractéristiques environnementales, etc.

Ces remarques décrivent des prolongements souhaitables des analyses. Ainsi, la disponibilité récente[23] de l'ensemble des données SHARE devrait permettre de contrôler l'état de santé à l'âge adulte avec d'autres informations, comme les comportements de santé individuels et de fournir un choix étendu de variables individuelles pouvant instrumenter la catégorie socioprofessionnelle. En ce qui concerne l'obtention de variables de meilleure qualité pour informer sur les conditions dans l'enfance, nous avons introduit un module de questions rétrospectives sur l'enfance dans l'Enquête Santé et Protection Sociale en 2006. Ces informations sur les habitudes de vie durant l'enfance, sur la santé ainsi que l'éducation des parents s'ajouteront donc aux riches données de santé individuelles contenues dans l'enquête.

Enfin, cette thèse apporte des éléments plus généraux à l'évaluation des politiques de réduction des inégalités. Nous pensons, par exemple, aux résultats des chapitres 4 et 5.

Dans le chapitre 4, la décomposition permet d'évaluer la contribution à l'inégalité de certaines variables et de ce fait, éclaire les groupes spécifiques sur lesquels les politiques doivent concentrer leurs efforts. En particulier, nos analyses suggèrent que la réduction de

[23] La version, dite *release 2*, de l'enquête SHARE a été mise en ligne au mois de septembre 2007.

l'association entre le revenu et la santé dans la société est plus favorable à la diminution des inégalités sociales de santé qu'à la diminution des inégalités de revenu elles-mêmes.

Alors que des analyses similaires appliquées aux inégalités sociales de recours aux soins ont permis de mettre en évidence les effets favorables de la mise en place de la couverture maladie universelle sur la consommation de soins, notre analyse ne permet pas de conclure que cette réforme est aussi à l'origine de la diminution de ces inégalités sociales de santé. Ce résultat n'est pas surprenant. Peu d'études ont pu établir l'impact d'une augmentation des consommations de soins sur l'état de santé à court terme. L'analyse sur données françaises de Jusot *et al.* (2005) met, par exemple, en évidence, à état de santé donné, qu'un accroissement du recours aux soins de généralistes a un impact limité sur l'état de santé quatre ans plus tard. En effet, les soins médicaux limiteraient seulement l'invalidité. Néanmoins, il est attendu que des améliorations s'observent sur l'état de santé de long terme et la réplication de cette analyse dans le temps permettra de faire la lumière sur ces aspects.

Par ailleurs, le chapitre 5 met en évidence l'existence d'autres leviers d'actions sur les inégalités sociales de santé : agir dès l'enfance. En France, il semble aller de soi qu'un élément d'action pour faire face aux disparités dans le domaine de la santé est de rendre l'accès aux soins plus équitable et donc d'intervenir dans le domaine de l'assurance. Or, notre analyse souligne la nécessité de contrecarrer les mauvais effets sur la santé d'un milieu social défavorable. Selon Deaton (2002), il s'agirait de mettre en place des politiques ciblées sur l'éducation. En effet, le rôle protecteur de l'éducation sur la santé a été mis en évidence empiriquement : une année supplémentaire d'éducation réduit les taux de mortalité à tout âge, d'environ 8% (Elo & Preston, 1996). De même dans nos analyses empiriques aux chapitres 4 et 5, nous observons de meilleurs états de santé pour les plus hauts niveaux d'éducation.

Cependant, quelle est la faisabilité d'une politique qui aurait ces objectifs ? Une politique agissant sur l'éducation s'envisage à long terme, voire très long terme et de fait, les décideurs politiques ont tendance à préférer des politiques aux répercussions immédiates ou de court terme sur la santé (Couffinhal *et al.*, 2005).

Même si les résultats du chapitre 4 suggèrent que les politiques redistributives sont à l'origine d'une diminution de l'élasticité de la santé avec le revenu et donc d'une baisse des inégalités, nous montrons au chapitre 4 comme au chapitre 5, que celles-ci ne suffisent pas à faire disparaître les inégalités sociales de santé.

Appendix A

Appendix related to chapter 4 and chapter 5

Kolmogorov-Smirnov test

The idea of standard tests of welfare dominance to compare distributions of welfare indicators is to make ordinal judgments. We can perform statistical inference on orderings using *Kolmogorov-Smirnov* tests.

Suppose that we have an i.i.d. sample $X_1, \ldots, X_n$ with some unknown distribution F and we would like to test the hypothesis that F is equal to a particular distribution F_0. The *Kolmogorov-Smirnov* test is designed to test a simple hypothesis $F = F_0$. In other words, we aim to decide between the following hypotheses:

$$H_0 : F = F_0, H_1 : F \neq F_0 \tag{A.1}$$

When more than two alternatives are considered, the test is performed for each pair of distributions.

The *Kolmogorov-Smirnov* test has the advantage to be distribution-free and non parametric. The KS-test is a robust test that cares only about the relative distribution of the data. The K-S test is based on the maximum distance between these two curves, which is called the D-statistic. It compares then this D-statistic against the critical D-statistic for that sample size. If the calculated D-statistic is greater than the critical one, then reject the null hypothesis that the distribution is of the expected form.

The Stata command ksmirnov has been used to perform the *Kolmogorov-Smirnov* test in this thesis.

Appendix B

Appendix related to chapter 5

B.1 Brant test for the use of ordered Logit regression models

The Brant test permits judging whether ordered logit regression is appropriate.

The test relies on four equations:

- Eq. 1: $SAH = 1$ compared to $SAH = 2, 3, 4, 5$
- Eq. 2: $SAH = 1, 2$ compared to $SAH = 3, 4, 5$
- Eq. 3: $SAH = 1, 2, 3$ compared to $SAH = 4, 5$
- Eq. 4: $SAH = 1, 2, 3, 4$ compared to $SAH = 5$

Our regression model contains 23 regressors and self-assessed health is described on five categories. Therefore, the Brant test relies on 96 tests. The higher the number of tests, the harder it is to confirm the assumption of parallel slopes. In other words, if there are many regressors, the assumption is likely to be violated. As a consequence, the significance level of the Brant test is often restricted to 1% (Long & Freese, 2006 ; Long, 1997).

The Stata command for Brant gives information.

- Estimated coefficients in binary regressions (Eq. 1: $y_1 > 1$,..., Eq. n: $y_{n-1} > n - 1$) as shown in table B.1.
- The global Wald test (cf. table B.2 permits judging whether at least one of the coefficients varies according to binary regressions.
- The Brant test displays distinctive tests for each regressor and allows identifying which variable breaks the assumption.

Estimated coefficients from j-1 binary regressions				
	y>1	y>2	y>3	y>4
Gender				
Woman	0,377	0,742	0,250	0,069
Man				
Age	-0,063	-0,041	-0,062	-0,073
Father's occupation				
Senior managers and professionals	-0,349	-0,141	0,084	0,355
Technicians and associate professionals and armed forces	0,284	-0,136	0,055	0,449
Office clerks and service workers and shop and market sales workers	-0,497	-0,479	0,166	0,496
Skilled agricultural and fishery workers	0,138	-0,026	0,204	0,419
Craftsmen and skilled workers	-0,625	-0,345	0,082	0,102
Elementary occupations and unskilled workers	ref.			
Mother's occupation				
Senior managers, professionals and technicians	-0,458	-0,459	0,038	-0,264
Office clerks and service workers and shop and market sales workers	0,395	0,605	0,232	-0,357
Skilled agricultural and fishery workers	-0,695	0,011	0,128	-0,042
Craft and related trades workers	1,225	0,059	-0,112	0,187
Elementary occupations and unskilled workers	-0,006	-0,081	-0,334	-0,528
Homemakers	ref.			
Father's relative longevity	0,009	0,004	0,007	0,005
Mother's relative longevity	-0,005	-0,007	0,008	0,017
Education level				
Elementary level diploma	-0,038	0,536	0,467	0,335
secondary level diplomas	0,115	0,821	0,539	0,330
Baccalauréat (A-levels)	-0,211	0,972	0,870	0,774
No diploma	ref.			
Descendants' occupation				
Senior managers and professionals	0,402	0,667	0,628	1,436
Technicians and associate professionals and armed forces	0,325	0,847	0,894	1,459
Office clerks and service workers and shop and market sales workers	0,165	0,612	0,431	0,997
Skilled agricultural and fishery workers	-0,040	0,822	0,251	0,082
Craftsmen and skilled workers	0,238	0,372	0,055	0,382
Elementary occupations and unskilled workers	ref.			
Homemakers	-0,337	-0,281	0,383	0,838
Cons	8,106	3,923	3,114	0,691

Table B.1: Brant test: binary regressions results (*2004/05 SHARE*).

Brant Test of Parallel regression assumption			
	chi2	p>chi2	df
All	77,12	0,235	69
Gender			
Woman**	9,68	0,022	3
Man	ref.		
Age**	8,09	0,044	3
Father's occupation			
Senior managers and professionals	0,81	0,846	3
Technicians and associate professionals and armed forces	1,17	0,761	3
Office clerks and service workers and shop and market sales workers	2,86	0,413	3
Skilled agricultural and fishery workers	0,67	0,881	3
Craftsmen and skilled workers	1,86	0,603	3
Elementary occupations and unskilled workers	ref.		
Mother's occupation			
Senior managers. professionals and technicians	3,6	0,308	3
Office clerks and service workers and shop and market sales workers	4,25	0,236	3
Skilled agricultural and fishery workers	1,95	0,584	3
Craft and related trades workers	2,41	0,491	3
Elementary occupations and unskilled workers	1,26	0,738	3
Homemakers	ref.		
Father's relative longevity	0,97	0,808	3
Mother's relative longevity**	10,03	0,018	3
Education level			
Elementary level diploma	2,59	0,46	3
secondary level diplomas	3,95	0,267	3
Baccalauréat (A-levels)	4,43	0,219	3
No diploma	ref.		
Descendants' occupation			
Senior managers and professionals	3,17	0,366	3
Technicians and associate professionals and armed forces	2,35	0,503	3
Office clerks and service workers and shop and market sales workers	2,53	0,47	3
Skilled agricultural and fishery workers	4,11	0,25	3
Craftsmen and skilled	ref.		
workers	1,71	0,635	3
Elementary occupations and unskilled workers	ref.		
Homemakers	5,78	0,123	3

Table B.2: Brant test: Wald Tests (*2004/05 SHARE*).

Significant test statistics provide evidence that the assumption of parallel slopes has been violated. Our results are paradoxical as the global Wald test is not significant and does not underline any violation whereas specific tests concerning each regressor emphasise three significant tests at 5%. Three variables break the assumption of parallel slopes, namely gender, age and mother's relative longevity (emphasised in table B.2 with **).

Nevertheless, if we apply a restricted level for significance thresholds then there is no longer violations of the assumption[1].

The solution is to apply a generalised ordered Logit as advised by Williams (2005). We carry out both a regression model under constraints, which reproduces the estimated parameters from the initial ordered Logit, i.e. imposing proportionality and a generalised ordered Logit, which relaxes this constraint[2].

In the unconstrained model, the estimated parameters changing the most are those associated to longevity, age and gender. We carry out a global test of the assumption by analysing how the two models differ.

lrtest constrained unconstrained	
Likelihood-ratio test	LR chi2(69) = 81.57
(Assumption: constrained nested in unconstrained)	Prob > chi2 = 0.1430

Table B.3: Differences between constrained and unconstrained models (*2004/05 SHARE*).

The Chi-2 statistic is similar to the one computed in the Brant test ($77,12$ in table B.2). Nevertheless, this statistic is more robust in this second test because the likelihood-ratio test is more robust than the Wald test. Considering that we again observe an unsignificant statistics, we can thus assume that the assumption is not violated by any variables.

[1]Nevertheless, this restriction appears excessive as estimated coefficients in binary regressions concerning these variables are qualitatively changing. In other words, the slope goes in the opposite direction according to the equation.

[2]Results are available on request

B.2 Inequalities in health in adulthood over parents' health

Variables	Father's longevity				Mother's longevity			
	Interval regression			CI	Interval regression			CI
	Coeff.	S.E	P-value		Coeff.	S.E	P-value	
Relative longevity	0,0003	0,0001	0,009	0,4064	0,0003	0,0001	0,043	0,5900
Female	0,0116	0,0044	0,008	0,0286	0,0120	0,0044	0,006	0,0030
Male	ref.							
49-54 years old	-0,0025	0,0085	0,766	-0,3282	-0,0031	0,0085	0,715	-0,5049
55-59 years old	ref.							
60-64 years old	-0,0134	0,0076	0,078	-0,1808	-0,0155	0,0076	0,043	-0,0927
65-69 years old	-0,0185	0,0073	0,011	-0,0995	-0,0207	0,0074	0,005	0,0493
70-74 years old	-0,0288	0,0072	0	-0,0234	-0,0305	0,0074	0	0,1044
75-79 years old	-0,0505	0,0077	0	0,3150	-0,0503	0,0078	0	0,2379
80-84 years old	-0,0627	0,0083	0	0,3543	-0,0622	0,0084	0	0,2694
>=85 years old	-0,0819	0,0107	0	0,2392	-0,0827	0,0108	0	0,2567
Elementary level diploma	0,0170	0,0054	0,002	0,0634	0,0170	0,0054	0,002	0,0438
Secondary level diplomas	0,0197	0,0060	0,001	-0,0742	0,0191	0,0060	0,001	-0,0139
Baccalauréat (A-levels)	0,0386	0,0075	0	0,0240	0,0387	0,0075	0	-0,0064
No diploma	ref.							
Senior managers and professionals	0,0324	0,0088	0	0,0655	0,0323	0,0088	0	0,0706
Technicians and associate professionals and armed forces	0,0397	0,0082	0	-0,0418	0,0393	0,0082	0	-0,0132
Office clerks and service workers and shop and market sales workers	0,0232	0,0075	0,002	-0,0093	0,0229	0,0075	0,002	-0,0294
Skilled agricultural and fishery workers	0,0189	0,0091	0,038	0,1519	0,0192	0,0091	0,035	0,1016
Craftsmen and skilled workers	0,0064	0,0076	0,404	-0,1058	0,0056	0,0076	0,462	-0,0360
Elementary occupations and unskilled workers	ref.							
Homemakers	0,0118	0,0094	0,207	0,1072	0,0112	0,0094	0,233	0,0542
cons	0,6399	0,0091	0		0,6447	0,0089	0	

Table B.4: Interval regression models and concentration indices of the distribution of health over parents' longevity (*2004/05 SHARE*).

B.3 Inequalities in health in adulthood over social status and social background

Variables	Individual's SES				Father's SES				Mother's SES			
	Interval regression			CI	Interval regression			CI	Interval regression			CI
	Coeff.	S.E	P-value		Coeff.	S.E	P-value		Coeff.	S.E	P-value	
Ordered SES	0,0043	0,0009	0	0,2726	0,0016	0,0014	0,249	0,2580	0,0019	0,0016	0,239	0,1827
Female	0,0142	0,0040	0	-0,0785	0,0119	0,0044	0,006	-0,0181	0,0120	0,0044	0,006	0,0022
Male	ref.											
49-54 years old	-0,0050	0,0085	0,559	0,0349	-0,0047	0,0085	0,579	0,0877	-0,0046	0,0085	0,59	0,0386
55-59 years old	ref.											
60-64 years old	-0,0135	0,0076	0,076	0,0902	-0,0140	0,0076	0,066	0,0103	-0,0141	0,0076	0,064	0,0879
65-69 years old	-0,0178	0,0073	0,014	0,0442	-0,0183	0,0073	0,012	-0,0145	-0,0182	0,0073	0,012	-0,0406
70-74 years old	-0,0278	0,0073	0	-0,0787	-0,0279	0,0072	0	-0,0221	-0,0278	0,0072	0	-0,0412
75-79 years old	-0,0461	0,00760	0	-0,0561	-0,0467	0,0076	0	-0,0121	-0,0466	0,0076	0	-0,0354
80-84 years old	-0,0573	0,0081	0	-0,0372	-0,0584	0,0081	0	0,0099	-0,0582	0,0081	0	-0,0079
>=85 years old	-0,0773	0,0106	0	-0,0458	-0,0786	0,0107	0	0,0033	-0,0791	0,0107	0	0,0478
Elementary level diploma	0,0175	0,0054	0,001	-0,1856	0,0173	0,0054	0,001	-0,0897	0,0175	0,0054	0,001	-0,0783
Secondary level diplomas	0,0213	0,0059	0	0,1267	0,0197	0,0060	0,001	-0,0433	0,0200	0,0060	0,001	-0,0139
Baccalauréat (A-levels)	0,0423	0,0072	0	0,5456	0,0379	0,0077	0	0,3749	0,0388	0,0075	0	0,2757
No diploma	ref.											
Senior managers and professionals					0,0314	0,0089	0	0,2934	0,0319	0,0088	0	0,2269
Technicians and associate professionals and armed forces					0,0385	0,0082	0	0,1053	0,0387	0,0082	0	0,1161
Office clerks and service workers and shop and market sales workers					0,0226	0,0075	0,002	-0,0669	0,0226	0,0075	0,002	0,0108
Skilled agricultural and fishery workers					0,0195	0,0091	0,032	0,1065	0,0204	0,0091	0,025	-0,3457
Craftsmen and skilled workers					0,0059	0,0076	0,439	-0,2149	0,0058	0,0076	0,449	-0,1084
Elementary occupations and unskilled workers					ref.							
Homemakers					0,0110	0,0094	0,241	-0,0029	0,0110	0,0094	0,24	0,0361
Cons	0,639652	0,0078604	0		0,6417	0,0095	0		0,6395	0,0103	0	

Table B.5: Interval regression models and concentration indices of the distribution of health over social status and social background (*2004/05 SHARE*).

B.4 Module of questions on retrospective reports in 2006 IRDES-HHIS

As regard to the absence of a large data sample gathering together family background and current social and health information, we have proposed a module of questions on retrospective reports, which has been included in the 2006 IRDES-HHIS survey, carried out in France. We present below this module.

Nous allons maintenant vous poser quelques questions sur la ou les personne(s) qui vous élevai(en)t lorsque vous aviez 12 ans, qu'il s'agisse de vos parents ou non.

ENQ : Pour permettre aux interviewés de se repérer dans le temps, n'hésitez pas à préciser que "12 ans" correspond à la fin de l'école primaire et aux premières années du collège. Pour les personnes plus âgées, cela correspond aussi à l'âge de l'obtention du certificat d'études primaires.

Q 1. Lorsque vous aviez 12 ans, quelle était la situation professionnelle principale de l'homme qui vous élevait (votre père, votre beau-père...) ?

1. Il travaillait (allez en Q 2)
2. Il était au chômage (allez en Q 3)
3. Il était retraité, retiré des affaires, préretraité (allez en Q 3)
4. Il était inactif (homme au foyer, invalide,...) (allez en Q 3)
5. Il était temporairement absent du foyer à cette époque (appelé sous les drapeaux, hospitalisation...) (allez en Q 3)
6. Il était décédé (allez en Q 9)
7. Il n'y avait pas d'homme qui vous élevait à cette époque (allez en Q 17)
8. Vous viviez alors dans un foyer de l'assistance publique
9. [nsp] (allez en Q 17)

Q 2. Quelle était alors sa profession principale ?

Il est nécessaire de donner un intitulé exact. Par exemple, ne pas indiquer "employé" mais "vendeur de...", ne pas indiquer "ouvrier" mais "mon- teur". Pour un fonctionnaire, indiquer le titre exact, par exemple "inspecteur de police" ou "professeur agrégé"

Si cet homme était retraité, au chômage, inactif, en longue maladie ou décédé, il faut indiquer la dernière profession qu'il ait occupée.

Q 3. Quelle était alors sa dernière profession principale avant d'être [retraité, retiré des affaires, préretraité / inactif / absent momentanément] ?

Il est nécessaire de donner un intitulé exact. Par exemple, ne pas indiquer "employé" mais "vendeur de...", ne pas indiquer "ouvrier" mais "monteur". Pour un fonctionnaire, indiquer le titre exact, par exemple "inspecteur de police" ou "professeur agrégé"

Q 4. Toujours lorsque vous aviez 12 ans, quel était son niveau d'études ?

Préciser qu'il s'agit du niveau d'études que la personne avait au moment où l'interviewé était âgé de 12 ans. Des études reprises par la suite, quand l'interviewé était adulte, ne sont pas prises en considération.

1. non scolarisé
2. maternelle, primaire, certificat d'études primaire
3. 1er cycle : 6ème, 5ème, 4ème, 3ème, technique, jusqu.au CAP, BEP
4. 2nd cycle : 2nde, 1ère, terminale, Baccalauréat technique, Baccalauréat
5. études supérieures au baccalauréat
6. autres, Précisez. . .

7. [nsp]
8. [ne se souvient pas]

Q 5. Jusqu'à vos 12 ans, avait-il connu des périodes d'inactivité professionnelle involontaires d'au moins 6 mois (maladie, chômage...) l'amenant par exemple à être présent au foyer dans la journée ?

Il est bien important de mettre en évidence que l'on parle d'inactivité professionnelle involontaire et que la précision "l'amenant par exemple à être présent au foyer dans la journée" n'est qu'un exemple qui permet de donner un repère dans les souvenirs. En aucun cas, un militaire de retour de mission par exemple n'est pris en compte ici.

1. Non, jamais
2. Oui
3. [nsp]
4. [ne se souvient pas]
5. [n.a jamais travaillé entre votre naissance et vos 12 ans]

Si Q 5 = 2 :

Q 5.1 Était-ce, selon vous, principalement lié à des problèmes de santé ?

Il peut s'agir d'un problème de santé même quand la période d'inactivité a été un évènement ponctuel et que la personne n'est pas restée inactive définitivement

1. Oui
2. Non
3. [nsp]

Q 6. Lorsque vous aviez 12 ans, son état de santé général était-il, selon vous...

Il s'agit bien de l'état de santé au moment où l'interviewé avait 12 ans.

1. ...Très bon
2. ...Bon
3. ...Moyen
4. ...Mauvais
5. ...Très mauvais
6. [nsp]

Q 7. En quelle année est-il né ?

Si ne sait pas en Q 7 :

Q 7.1 En quelle année ou en quelle décennie est-il né approximativement ?

Q 8. Est-il toujours vivant ?

1. Oui (allez en Q 17)
2. Non
3. [nsp] (allez en Q 17)

Q 9. Quelle est l'année de son décès ?

Si ne sait pas en Q 9 :

Q 9.1 En quelle année ou en quelle décennie est-il décédé approximativement ?

Q 10. Quelle est la cause de son décès ?

Q 11. En quelle année était-il né ?

Q 12. En quelle année ou en quelle décennie était-il né approximativement ?

Q 13. Quelle était alors la dernière profession principale qu'il ait occupée avant son décès ?

Q 14. Quel était son niveau d'études ?

1. non scolarisé
2. maternelle, primaire, certificat d'études primaire

3. 1er cycle : 6ème, 5ème, 4ème, 3ème, technique, jusqu'au CAP, BEP
4. 2nd cycle : 2nde, 1ère, terminale, Baccalauréat technique, Baccalauréat
5. études supérieures au baccalauréat
6. autres. Précisez
7. [nsp]
8. [ne se souvient pas]

Q 15. Avant son décès, avait-il connu des périodes d'inactivité professionnelle involontaire d'au moins 6 mois (maladie, chômage...) l'amenant par exemple à être présent au foyer dans la journée ?

Il est important de mettre en évidence que l'on parle d'inactivité professionnelle involontaire et que la précision l'amenant par exemple à être présent au foyer dans la journée n'est qu'un exemple qui permet de donner un repère dans les souvenirs. En aucun cas, un militaire de retour de mission par exemple n'est pris en compte ici.

1. Non, jamais (allez en Q 17)
2. Oui, au moins une fois
3. Oui, à plusieurs reprises
4. [nsp] (allez en Q 17)
5. [ne se souvient pas] (allez en Q 17)

Q 16. Était-ce, selon vous, principalement lié à des problèmes de santé ? Il peut s'agir d'un problème de santé même quand la période d'inactivité a été un évènement ponctuel et que la personne n'est pas restée inactive définitivement.

1. Oui
2. Non
3. [nsp]

Q 17. Lorsque vous aviez 12 ans, quelle était la situation professionnelle principale de la femme qui vous élevait (votre mère, votre belle-mère...) ?

1. elle travaillait
2. elle était au chômage (allez en Q 19)
3. elle était retraitée, retirée des affaires, préretraitée (allez en Q 19)
4. elle était inactive (femme au foyer, invalide...) (allez en Q 19)
5. elle était absente du foyer à cette époque (hospitalisation. . .) (allez en Q 19)
6. elle était décédée (allez en Q 24)
7. Il n'y avait pas de femme qui vous élevait à cette époque (allez en Q 30)
8. [nsp]

Q 18. Quelle était alors sa profession principale ?

Il est nécessaire de donner un intitulé exact. Utiliser "Femme au foyer" uniquement pour les personnes n'ayant jamais travaillé.

Q 19. Quelle était alors sa dernière profession principale avant d'être [retraitée, retirée des affaires, préretraitée / inactive / absente momentanément] ?

Q 20. Toujours lorsque vous aviez 12 ans, quel était son niveau d'études ?

1. non scolarisé
2. maternelle, primaire, certificat d'études primaire
3. 1er cycle : 6ème, 5ème, 4ème, 3ème, technique, jusqu.au CAP, BEP
4. 2nd cycle : 2nde, 1ère, terminale, Baccalauréat technique, Baccalauréat
5. études supérieures au baccalauréat
6. autres. Précisez. . .
7. [nsp]

8. [ne se souvient pas]

Q 21. Lorsque vous aviez 12 ans, son état de santé général était-il, selon vous...

1. ...Très bon
2. ...Bon
3. ...Moyen
4. ...Mauvais
5. ...Très mauvais
6. [nsp]

Q 22. En quelle année est-elle née ?

Si ne sait pas en Q 22 :

Q 22.1 En quelle année ou en quelle décennie est-elle née approximativement ?

Q 23. Est-elle toujours vivante ?

1. Oui (allez en Q 30)
2. Non
3. [nsp] (allez en Q 30)

Q 24. Quelle est l'année de son décès ?

Si ne sait pas en Q 24 :

Q 24.1 En quelle année ou en quelle décennie est-elle décédée approximativement ?

Q 25. Quelle est la cause de son décès ?

Q 26. En quelle année était-elle née ?

Q 27. En quelle année ou en quelle décennie est-elle née approximativement ?

Q 28. Quelle était alors la dernière profession principale qu'elle ait occupée avant son décès ?

Q 29. Quel était son niveau d'études ?

1. non scolarisé
2. maternelle, primaire, certificat d'études primaire
3. 1er cycle : 6ème, 5ème, 4ème, 3ème, technique, jusqu'au CAP, BEP
4. 2nd cycle : 2nde, 1ère, terminale, Baccalauréat technique, Baccalauréat
5. études supérieures au baccalauréat
6. autres. Précisez
7. [nsp]
8. [ne se souvient pas]

Je vais maintenant vous poser des questions plus générales concernant votre enfance et le foyer dans lequel vous avez grandi.

Q 30. Lorsque vous aviez 12 ans, diriez-vous que la ou les personne(s) qui vous élevai(en)t (parents, beaux-parents,...) étaient financièrement...

ENQ : CITER - une seule réponse

1. ...très à l'aise
2. ...plutôt à l'aise
3. ...plutôt gênés
4. ...très gênés
5. [nsp]

Q 31. Si vous comparez votre niveau de vie à celui des (ou de la) personne(s) qui vous élevaient, lorsqu'elles avaient [âge calculé] ans, c'est-à-dire le même âge que vous aujourd'hui, diriez-vous que votre niveau de vie est aujourd'hui :

1. BIEN MEILLEUR que le leur à cette époque
2. MEILLEUR que le leur à cette époque
3. IDENTIQUE au leur à cette époque
4. INFERIEUR au leur à cette époque
5. BIEN INFERIEUR au leur à cette époque
6. elles (elle) étai(en)t décédée(s) à cet âge-là
7. [nsp] Précisez
8. [refus] Précisez

Q 32. Lorsque vous aviez 12 ans, y avait-il une personne qui fumait dans le foyer où vous avez grandi ?

ENQ : CITER - plusieurs réponses possibles

1. Oui, l'homme qui vous élevait
2. Oui, la femme qui vous élevait
3. Oui, vous-même
4. Oui, une autre personne (soeur, frère,...)
5. Non, personne ne fumait
6. [nsp]

Q 33. Durant votre enfance, pensez-vous qu'une personne avec qui vous avez vécu ait eu un problème d'alcool ?

ENQ : CITER - plusieurs réponses possibles. La question se pose sur l'enfance en général et non pas seulement à l'âge de 12 ans. Il est donc possible qu'il n'y ait plus d'homme ou de femme dans le foyer de l'interviewé lorsqu'il a 12 ans, mais que par le passé, il y ait bien eu un homme ou une femme qui l'élevait et qui avait un problème avec l'alcool et cette information est importante.

1. Oui, l'homme qui vous a élevé(e)
2. Oui, la femme qui vous a élevé(e)
3. Oui, vous-même
4. Oui, une autre personne (soeur, frère, beau-père, belle-mère...)
5. Non, personne n'avait de problème avec l'alcool
6. [nsp]

References

1. Abul Naga, R. H. & Yalcin, T. 2007. Inequality measurement for ordered response health data. Working Paper of the Distributional Analysis Research Programme,(92).

2. Allison, P. D. 1999. *Logistic regression using the SAS System: theory and application*. Cary: SAS Institute Press.

3. Allison, R. A. & Foster, J. E. 2004. Measuring health inequality using qualitative data. *Journal of Health Economics*, 23(3): 505-524.

4. Anderson, K. H. & Burkhauser, R. V. 1985. The retirement-health nexus: a new measure of an old puzzle. *Journal of Human Resources*, 20: 315-330.

5. Apouey, B. 2007. Measuring health polarization with self-assessed health data. *Health Economics*, 16: 875-894.

6. Arneson, R. J. 1989. Equality and equal opportunity of welfare. *Philosophical Studies*, 56: 77-93.

7. Atkinson, T. 1970. On the measurement of inequality. *Journal of Economic Theory*, 2: 244-263.

8. Atkinson, T. & Bourguignon, F. 1987. Income distribution and differences in needs. In G. R. Feiwel (Ed.), *Arrow and the foundation of the theory of economic policy*: 350-370. London.

9. Atkinson, T. & Bourguignon, F. 1982. The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies*, XLIX: 183-201.

10. Atkinson, T. Cantillon, B. Marlier, E. & Nolan, B. 2001. Indicators for social inclusion in the European Union. Actes de la Conférence de l'Union Européenne, Anvers, 14-15 septembre

11. Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., O'Donnell, O., & Chatterji, S. 2006. Does heterogeneity bias the measurement of health? Tinbergen Institute Discussion Papers, No 06-033/3.

12. Baker, M., Stabile, M., & Deri, C. 2004. What do self-reported, objective, measures of health measure? *Journal of Human Resources*, 39(4): 1067-1093.

13. Barker, D. J. P. 1997. Fetal growth and cardiovascular disease in later life. In M. Marmot & M. E. J. Wadsworth (Eds.), *Fetal and early childhood environment*: 96-108.

14. Barker, D. J. P. 1996. Fetal origins of coronary heart disease. *Britain Medicine Journal*, 311: 171-174.

15. Barker, D. J. P., Forsen, T., Uutela, A., Osmond, C., & Eriksson, G. J. 2001. Size at birth and resilience to effects of poor living conditions in adult life: longitudinal study. *Britain Medicine Journal*, 323: 1273-1276.

16. Barnay, T., Jusot, F., Rochereau, T., & Sermet, C. 2005. Les mesures de la santé et de l'activité sont-elles comparables dans les enquêtes européennes ? *Questions D'Économie De La Santé*, IRDES, (96).

17. Bartel, A. & Taubman, P. 1979. Health and labor market success: the role of various diseases. *Review of Economics and Statistics*, 61(1): 1-8.

18. Bartley, M., Power, C., Blane, D., & Davey Smith, G. 1994. Birth weight and later socioeconomic disadvantage: evidence from the 1958 British cohort study. *Britain Medicine Journal*, 309: 1475-1478.

19. Belcher, D., Wurapa, F., Neumann, A., & Lourie, I. 1976. Comparison of morbidity interviews with a health examination survey in rural Africa. American *Journal of Tropical Medicine and Hygiene*, 25: 751-758.

20. Bellanger, M. & Jourdain, A. 2004. Tackling regional inequalities in France by resource allocation: a case for complementary instrumental and process-based approaches. *Applied Health Econometrics and Health Policy*, 3(4): 243-250.

21. Benyamini, Y., Leventhal, E. A., & Leventhal, H. 2000. Gender differences in processing information for making self-assessments of health. *Psychosomatic Medicine*, 62: 354-364.

22. Bergson, H. 1920. *L'évolution créatrice*. Paris: Librairie Félix Alcan.

23. Berkman, L. F. & Kawachi, I. 2000. *Social epidemiology*. Oxford University Press. New York.

24. Blane, D., Hart, C. L., & Davey Smith, G. 1996. Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *Britain Medicine Journal*, 313: 1434-1438.

25. Blau, D. M. & Gilleskie, D. B. 2001. The effect of health on employment transitions of older men. Worker Wellbeing in a Changing Labor Market: 35-65. JAI Press.

26. Blaxter, M. 1990. *Health and lifestyles*. London and New York: Tavistock/Routledge Publication.

27. Bleichrodt, H. & van Doorslaer, E. 2006. A welfare economics foundation for health inequality measurement. *Journal of Health Economics*, 25(5): 945-957.

28. Boisguérin, B. 2005. Les bénéficiaires de la CMU au 31 décembre 2003. *Etudes Et Résultats*, DREES, 381.

29. Boissonnat, V. & Mormiche, P. 2007. Handicaps et inégalités sociales en France, 1999. *Bulletin Épidémiologique Hebdomadaire*, 2-3: 26-28.

30. Bolin, K., Jacobson, L., & Lindgren, B. 2002. The family as the health producer - when spouses act strategically. *Journal of Health Economics*, 21: 475-495.

31. Bolin, K., Jacobson, L., & Lindgren, B. 2001. The family as the health producer - when spouses are Nash-bargainers. *Journal of Health Economics*, 20(362): 349.

32. Bommier, A. & Stecklov, G. 2002. Defining health inequality : why Rawls succeeds where social welfare theory fails? *Journal of Health Economics*, 21(3): 497-513.

33. Börsch-Supan, A., Brugiavini, A., Jürges, H., Mackenbach, J. P., Siegrist, J., & Weber, G. 2005. Health, ageing and retirement in Europe - First results from the Survey of Health, Ageing and Retirement in Europe. Mannheim. Report.

34. Börsch-Supan, A. Jürges, H. & Lipps, O. 2003. SHARE: Building a panel survey on health, ageing and retirement in Europe. Report 32. Manheimer Forschunginstitut Okonomie und Demographischer Wandel.

35. Bound, J. 1990. Self-reported versus objective measures of health in retirement models. The *Journal of Human Resources*, 26(1): 107-137.

36. Bound, J., Schoenbaum, M., Stinebrickner, T. R., & Waidmann, T. 1999. The dynamics effects of health on the labor force transitions of older workers. *Labour Economics*, 6(2): 179-202.

37. Braveman, P., Krieger, N., & Lynch, J. 2000. Health inequalities and social inequalities in health. *Bulletin of the World Health Organisation*, 78(2): 232-235.

38. Brazier, J., Roberts, J., & Deverill, M. 2002. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21: 271-292.

39. Brazier, J., Tsuchiya, A., & Roberts, J. 2004. A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Economics*, 13(90): 873-884.

40. Brazier, J., Usherwood, T., Harper, R., & Thomas, K. 1998. Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology*, 51(11): 1115-1128.

41. Breen, R. 2005. *Social mobility in Europe*. Oxford University Press.

42. Brown, M. C. 1994. Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data. *Social Science and Medicine*, 38(9): 1243-1256.

43. Bryant, L. L., Beck, A., & Fairclough, D. L. 2000. Factors that contribute to positive perceived health in older population. *Journal of Ageing and Health*, 12(2): 169-192.

44. Burström, B. 2002. Increasing inequalities in health care utilisation across income groups in Sweden during the 1990s? *Health Policy*, 62(2): 117-130.

45. Burström, B. & Fredlund, P. 2001. Self-rated health: is it as good predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Health*, 55: 836-840. 46. Butler, J. S., Burkhauser, R. V.,

Mitchel, J. M., & Pincus, T. P. 1987. Measurement error on self-reported health variables. *The Review of Economics and Statistics*, 69(4): 644-650.

47. Cambois, E. & Jusot, F. 2007. Ampleur, tendance et causes des inégalités sociales de santé et de mortalité en Europe: une revue des études comparatives. *Bulletin Épidémiologique Hebdomadaire*, 2-3(janvier): 10-15.

48. Case, A., Fertig, A., & Paxson, C. 2005. The lasting impact of childhood health and circumstance. *Journal of Health Economics*, 24: 365-389.

49. Case, A., Lubotsky, D., & Paxson, C. 2002. Economic status and health in childhood: the origins of the gradient. *American Economic Review*, 92(5): 1308-1334.

50. Chambaz, C. & Maurin, E. 1998. Atkinson and Bourguignon's dominance criteria: extended and applied to the measurement of poverty in France. *Review of Income and Wealth*, 44(4): 497-513.

51. Chauvin, P. & Lebas, J. 2007. Inégalités et disparités sociales de santé en France. In : *Etat de santé et principales pathologies*. Chapitre 42. 331-341.

52. Chenu, A. 2000. Le repérage de la situation sociale. In A. Leclerc, D. Fassin, H. Grandjean, M. Kaminski, & T. Lang (Eds.), *Les inégalités sociales de santé*: 93-107. Paris.

53. Clark, A. E. & Vicard, A. 2007. Conditions de collecte et santé subjective : analyse sur données européennes. Documents De Travail PSE, WP nř 2007-10.

54. Clarke, P., Gerdtham, U. G., Johannesson, M., Bingefors, K., & Smith, L. 2002. On the measurement of relative and absolute income-related health inequality. *Social Science and Medicine*, 55(11): 1923-1928.

55. Cleary, P. D. & Jette, A. M. 2000. Reliability and validity of the functional status questionnaire. *Quality of Life Research*, 9(1): 747-753.

56. Com-Ruelle, L., Dourgnon, P., Jusot, F., Latil, E., & Lengagne, P. 2006. Identification et mesure des problèmes d'alcool en France : une comparaison de deux enquêtes en population générale. Rapport IRDES, 1600(9).

57. Com-Ruelle, L. Lecomte, T. Le Fur, P. Mizrahi, An. Mizrahi, Ar. & Sermet, C. 1997. Un indicateur de gravité des maladies: enquête sur la santé et les soins médicaux, France 1991-92. Rapport 1189. IRDES.

58. Couffinhal, A. 1999. Concurrence en assurance santé: entre efficacité et sélection. Université Paris Dauphine.

59. Couffinhal, A., Dourgnon, P., Geoffard, P.-Y., Grignon, M., Jusot, F., Lavis, J. N., Naudin, F., & Polton, D. 2005. Politiques de réduction des inégalités de santé, quelle place pour le système de santé ? Un éclairage européen. *Questions D'Économie De La Santé*, IRDES, 93.

60. Couffinhal, A., Dourgnon, P., & Tubeuf, S. 2004. Outils de mesure des inégalités: quelques débats d'actualité. *Santé, Société Et Solidarité*, 2004(2): 163-171.

61. Couffinhal, A. & Paris, V. 2003. Cost sharing in France. Study for the European Commission coordinated by the European Observatory on Health Care Systems.

62. Cowell, F. A. 2000. Measurement of inequality. In T. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution*: 89-169. London: Elsevier Science.

63. Cribier, F. 1997. L'allongement de la vie et la mortalité différentielle des travailleurs après la retraite, l'étude d'une cohorte de Parisiens. *Sociologie Et Santé*, 16(juin): 20-39.

64. Crossley, T. F. & Kennedy, S. 2002. The reliability of self-assessed health status. *Journal of Health Economics*, 21: 643-658.

65. Cutler, D. M. & Richardson, E. 1997. Measuring the health of the United States population. *Microeconomics*, Brookings Paper in Economic Activity: 217-271.

66. Daguet, Fabienne. Un siècle de fécondité française, caractéristiques et évolution de la fécondité de 1901 à 1999. *Insee Résultats* Société[8]. 2002.

67. Dalton, H. 1920. The measurement of inequality of incomes. *The Economic Journal*, 30(119): 348-361.

68. Dauphinot, V., Naudin, F., Guéguen, R., Perronnin, M., & Sermet, C. 2006. Ecarts entre morbidité déclarée et morbidité diagnostiquée - L'exemple de l'obésité, de l'hypertension artérielle et de l'hypercholestérolémie. *Questions D'Économie De La Santé*, IRDES,(114).

69. de Koninck, M. & Fassin, D. 2004. Les inégalités sociales de santé, encore et toujours. *Santé, Société Et Solidarité*, 2(Editorial): 4-12.

70. Deaton, A. 2001. Inequalities in income and inequalities in health. In F. Welch (Ed.), *The causes and consequences of increasing inequality*: Chicago.

71. Deaton, A. 2002. Policy implications of the gradient of health and wealth. *Health Affairs*, 21(2): 13-30.

72. Deaton, A. 2003. Health, income and inequality. NBER Reporter, Research Summary(Spring).

73. DeSalvo, K., Fan, V. S., McDonnell, M. B., & Fihn, S. D. 2005. Predicting mortality and healthcare utilization with a single question. *Health Services Research*, 40(4): 1234-1246.

74. Desplanques, G. 1993. L'inégalité sociale devant la mort. In: *La société française : données sociales 1993*: Paris.

75. Desplanques, Guy. La mortalité des adultes suivant le milieu social 1955-1971. *Collection de l'INSEE* - série D 195[44]. 1976.

76. Desplanques, G. 1984. L'inégalité sociale devant la mort. *Economie Et Statistique*, (162): 29-50.

77. Dezateux, C. & Stocks, J. 1998. Lung development and early origins of childhood respiratory illness. In M. Marmot & M. E. J. Wadsworth (Eds.), *Fetal and early childhood environment*: 40-57.

78. Dias, P. R. & Jones, A. 2007. Giving equality of opportunity a fair innings. *Health Economics*, 16: 109-112.

79. Disney, R., Emmerson, C., & Wakefield, M. 2004. Ill health and retirement in Britain: a panel data-based analysis. Unpublished Work, Presented at the ECuity III Workshop in Bonn - sept 2005.

80. Donaldson, C., Jones, A. M., Mapp, T., & Olson, J. A. 1998. Limited dependent variables in willingness to pay studies: applications in health care. *Applied Econometrics*, 30: 667-677.

81. Dourgnon, P., Grignon, M., & Jusot, F. 2001. L'assurance maladie réduit-elle les inégalités sociales de santé ? *Questions D'Économie De La Santé*, IRDES, 43.

82. Dube, S. R., Williamson, D. F., Thompson, T., Felitti, V. J., & Anda, R. F. 2004. Assessing the reliability of retrospective reports of adverse childhood experiences among adult HMO members attending a primary care clinic. *Child Abuse & Neglect*, 28(7): 729-737.

83. Dworkin, R. 1981. What is equality ? Part I: Equality of Welfare. *Philosophy and Public Affairs*, 10: 185-246.

84. Elo, I. & Preston, S. H. 1996. Educational differentials in mortality: United States, 1979-85. *Social Science and Medicine*, 42(1): 47-57.

85. Erreygers, G. 2006. Beyond the health concentration index: an Atkinson alternative for the measurement of the socioeconomic inequality of health. Unpublished Work, Research Paper 2006-029 - Presented at the Seminar EconomiX - Université Paris-Nanterre.

86. Etilé, F. & Milcent, C. 2006. Income-related reporting heterogeneity in self-assessed health: evidence from France. *Health Economics*, 15(9): 965-981.

87. Farr, W. 1839. Second Report. In: *Annual report of the Registrar-General of Births, Deaths and Marriages in England*: London.

88. Feeny, D. H., Torrance, G. W., & Furlong, W. J. 1996. Health utilies index. In B. Spilker (Ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials*: 239-252. Philadelphia.

89. Feinstein, L., Sabates, R., Anderson, T. M., Sorhaindo, A., & Hammond, C. 2006. What are the effects of education on health? In R. Desjardins & T. Schuller (Eds.), *Measuring the effects of education on health and civic engagement* - Proceedings of the Copenhagen Symposium: 171-313.

90. Ferraro, K. F. & Farmer, M. M. 1999. Utility of health data from social surveys: Is there a gold standard for measuring morbidity. *American Sociological Review*, 64(2): 303-315.

91. Ferrer-I-Carbonell, A. & Frijters, P. 2004. How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, 114(497): 641-659.

92. Fleurbaey, M. 2005. Health, wealth and fairness. *Journal of Public Economic Theory*, 7: 253-284.

93. Fleurbaey, M. 2007. Le revenu équivalent-santé, un outil pour l'analyse des inégalités sociales de santé. *Revue D'Epidémiologique Et De Santé Publique*, 55(1).

94. Fleurbaey, M. 1995. Equality and responsability. *European Economic Review*, 39: 683-689.

95. Fleurbaey, M. & Schokkaert, E. 2007. Unfair inequalities in health and health care. Unpublished Work, Presented at the 6th iHEA World Congress - Copenhagen 2007.

96. Fonseca, L. R. & Jones, A. 2007. Inequalities in self-assessed health in the Health Survey of England. Unpublished Work, ECuity III Working Paper 11.

97. Foster, J. E. & Sen, A. 1997. On economic inequality after a quarter century. *On economic inequality*: Oxford: Oxford University Press.

98. Foster, J. E. & Shorrocks, A. 1988. Poverty orderings and welfare dominance. *Social Choice and Welfare*, 5: 179-198.

99. Francesconi, C., Perronnin, M., & Rochereau, T. 2006. Complémentaire maladie d'entreprises: contrats obligatoires ou facultatifs, lutte contre l'antisélection et conséquences pour les salariés. *Questions D'Économie De La Santé*, IRDES,(115).

100. Franks, P., Gold, M. R., & Fiscella, K. 2003. Sociodemographics, self-rated health, and mortality in the US. *Social Science and Medicine*, 56(12): 2505-2514.

101. Fressin, A. 1967. *La perception chez Bergson et Merleau-Ponty*. Paris: Société d'Edition d'Enseignement Supérieur.

102. Fulghum, R. 1986. *All I really need to know I learned in kindergarten: uncommon thoughts on common things*. New York.

103. Fylkesnes, K. & Forde, O. H. 1992. Determinants and dimensions in self-evaluation of health. *Social Science and Medicine*, 35(3): 271-279.

104. Gakidou, E., Murray, C. J. L., & Frenk, J. 2000. Defining and measuring health inequality: an approach based on the distribution of health expectancy. *Bulletin of the World Health Organisation*, 78(1): 42-54.

105. Garcia Gomez, P. & Lopez, A.N. 2007. Regional differences in socio-economic health inequalities in Spain 1973/74 - 2003. Fundacion BBVA.

106. Gerdtham, U. G., Johannesson, M., Lundberg, L., & Isacson, D. 1999. A note on validating Wagstaff and van Doorslaer's health measure inequalities in health. *Journal of Health Economics*, 18(1): 117-124.

107. Girard, F., Cohidon, C., & Briançon, S. 2000. Les indicateurs globaux de santé. In A. Leclerc, D. Fassin, H. Grandjean, M. Kaminski, & T. Lang (Eds.), *Les inégalités sociales de santé*.

108. Goldberg, M. 1997. Surveillance épidémiologique des risques professionnels : l'exemple de l'amiante. *Revue Française Des Affaires Sociales*, 51(81): 89.

109. Goldberg, M., Melchior, M., Leclerc, A., & Lert, F. 2002. Les déterminants sociaux de la santé: apports récents de l'épidémiologie sociale et des sciences sociales de la santé. *Sciences Sociales Et Santé*, 20(4): 75-128.

110. Gravelle, H. 2003. Measuring income related inequality in health: standardisation and the partial concentration index. *Health Economics*, 12: 803-820.

111. Gravelle, H. & Sutton, M. 2003. Income-related inequalities in self-assessed health in Britain: 1979-1995. *Journal of Epidemiology and Community Health*, 57: 125-129.

112. Grignon, M. 2003. Etude sur un mécanisme de type APS (Allocation Personnalisée Santé) remplacant la CMU. Rapport réalisé pour la Fédération Interprofessionnelle des Mutuelles. IRDES.

113. Grignon, M. & Perronnin, M. 2003. Impact de la couverture maladie universelle complémentaire sur les consommations de soins. *Questions D'Économie De La Santé*, IRDES, 74 (2003).

114. Grignon, M., Perronnin, M., & Lavis, J. N. 2007. Does free complementary health insurance help the poor to access health care? Evidence from France. *Health Economics*, in Press (Published on line June 2007).

115. Groot, W. 2000. Adaptation and scale of reference bias in self-assessments of quality of life. *Journal of Health Economics*,(19): 403-420.

116. Grossman, M. 1972. The demand for health: a theoretical and empirical investigation. New York.

117. Gwatkin, D. Davidson, R. Rutstein, S. Johnson, K., Pande, R.P. & Wagstaff, A. 2000. Socio-economic differences in health, nutrition, and population in Indonesia. Health Nutrition, and Population, World Bank.

118. Hadar, J. & Russell, W. R. 1974. Stochastic dominance in choice under uncertainty. In D. L. Balch, D. MacFadden, & S. Y. Wu (Eds.), *Essays on Economic Behaviours and Uncertainty*.

119. Hardy, G. H., Littlewood, J. E., & Polya, G. 1952. *Inequalities*. Cambridge.

120. Haveman, R. H. & Wolfe, B. L. 1984. The decline in male labor force participation: a comment. *Journal of Political Economy*, 92(June 1984): 532-541.

121. Hernandez-Quevedo, C., Jones, A., & Rice, N. 2005. Reporting bias and heterogeneity in self-assessed health: evidence from the British Household Panel Survey. Unpublished Work, ECuity III WP 19.

122. Herpin, N. 2003. La taille des hommes: son incidence sur la vie en couple et la carrière professionnelle. *Economie Et Statistique*, 361: 71-90.

123. Huber, H. 2006. Vieillissement, dépenses de santé et inégalités de recours aux soins: essais de micro-économétrie appliquée. Université Paris X Nanterre.

124. Humphries, K. H. & van Doorslaer, E. 2000. Income-related health inequality in Canada. *Social Science and Medicine*, 50: 663-671.

125. Hyde, M., Jakub, H., Melchior, M., Van Oort, F., & Weyers, S. 2006. Comparison of the effects of low childhood socioeconomic position and low adulthood socioeconomic position on self-rated health in four European studies. *J Epidemiol Community Health*, 60: 882-886.

126. Idler, E. L. & Benyamini, Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behaviour*, 38(March): 21-37.

127. Idler, E. L. & Kasl, S. V. 1995. Self-ratings of health: do they also predict change in functional disability? *J Gerontol, Soc Sci*(50B): 344-353.

128. Idler, E. L. & Kasl, S. V. 1991. Health perceptions and survival: do global evaluations of health status really predict mortality? *J Gerontol, Soc Sci*(46): 55-65.

129. Jacobson, L. 2000. The family as a producer of health - An extended Grossman model. *Journal of Health Economics*, 19(5): 611-637.

130. Jenkins, S. P. & Lambert, P. J. 1993. Ranking income distributions when needs differ. *Review of Income and Wealth*, 39(4): 337-356.

131. Jones, A. M. & Lopez, A. N. 2004. Measurement and explanation of socioeconomic inequality in health with longitudinal data. *Health Economics*, 13: 1015-1030.

132. Jones, A. M., Rice, N., Bago d'Uva, T., & Balia, S. 2007b. Inequality in health utility and self-assessed health. *Applied Health Economics*: 29-50. Routledge Advanced Texts in Economics and Finance. London and New York.

133. Jones, A. M., Rice, N., Bago d'Uva, T., & Balia, S. 2007a. *Applied health economics*. Routledge Advanced Texts in Economics and Finance. London and New York.

134. Jougla, E., Rican, S., Péquignot, F., & Le Toullec, A. 2000. La mortalité. In A. Leclerc, D. Fassin, H. Grandjean, M. Kaminski, & T. Lang (Eds.), *Les inégalités sociales de santé*.

135. Juniper, E. F., Guyatt, G. H., Ferrie, P. J., & Griffith, L. E. 1993. Measuring quality of life in asthma. *American Review of Respiratory Disease*, 147: 832-838.

136. Jürges, H. 2007. True health vs response styles: exploring cross-country differences in self-reported health. *Health Economics*, 16(2): 163-178.

137. Jusot, F. 2003. Revenu et mortalité: analyse économique des inégalités sociales de santé en France. Ecole des Hautes Etudes en Sciences Sociales - DELTA.

138. Jusot, F., Couralet, P.-E., & Lengagne, P. 2005. Effets cumulatifs à long terme de l'accès aux soins et du revenu sur les inégalités de santé. Rapport de l'appel à projets de recherche ”Construction et représentation des inégalités sociales” de la MiRe. Paris. IRDES.

139. Kakwani, N., Wagstaff, A., & van Doorslaer, E. 1997. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics*, 77: 87-103.

140. Kaplan, G., Barell, V., & Lusky, A. 1988. Subjective state of health and survival on elderly adults. *Journal of Gerontology*, 43: 114-120.

141. Kapteyn, A., Smith, J. P., & van Soest, A. 2007. Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, Forthcoming.

142. Kerkhofs, M. & Lindeboom, M. 1995. Subjective health measures and state dependent reporting errors. *Health Economics*, 4: 221-235.

143. Khlat, M. & Sermet, C. 2004. La santé des chômeurs en France: revue de la litterature. *Revue D'Epidémiologique Et De Santé Publique*, 52(3): 465-474.

144. King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. 2003. Enhancing the validity and cross-cultural comparability of survey research. *American Political Science Review*, 97(4).

145. Knäuper, B. & Turner, P. A. 2003. Measuring health: improving the validity of health assessments. *Quality of Life Research*, 12: 81-89.

146. Kolm, S.-C. 1969. The optimal production of social justice. In H. Guitton & J. Margolis (Eds.), *Economie Publique*: Paris.

147. Koolman, X. & van Doorslaer, E. 2004. On the interpretation of a concentration index of inequality. *Health Economics*, 13(7): 649-656.

148. Krueger, D. E. 1957. Measurement of prevalence of chronic disease by household interviews and clinical evaluation. *American Journal of Public Health*, 47: 953-960.

149. Kuhn, R., Rahman, O., & Menken, J. 2004. Relating self-reported and objective health indicators to adult mortality in Bangladesh. *Demography*, under review.

150. Kunst, A. E., Groenhof, F., Mackenbach, J. P., & the EU Working Group on Socioeconomic Inequalities in Health 2000. Inégalités sociales de mortalité prématurée : la France comparée aux autres pays européens. In A. Leclerc, D. Fassin, H. Grandjean, M. Kaminski, & T. Lang (Eds.), *Les inégalités sociales de santé*: 53-68. La Découverte.

151. Kunst, A. E. & Mackenbach, J.P. 1996. La mesure des inégalités de santé d'origine socio-économique. Organisation mondiale de la santé, Bureau régional de l'Europe. Copenhague.

152. Kunst, A. E. & Mackenbach, J. P. 1994. The size of mortality differences associated with educational level in nine industrialised countries. *Public Health*, 84(6).

153. Landais, C. 2007. Les hauts revenus en France (1998-2006): Une explosion des inégalités? Documents De Travail PSE.

154. Lauridsen, J., Christiansen, T., & Häkkinen, U. 2004. Measuring inequality in self-reported health - discussion of a recently suggested approach using Finnish data. *Health Economics*, 13(7): 725-732.

155. Le Breton, M. 1986. Essais sur les fondements de l'analyse économique de l'inégalité. Université de Rennes I.

156. Le Galès, C., Costet, N., Gentet, J.-C., Kalifa, C., Frappaz, D., Edan, C., Sariban, E., Plantaz, D., & Doz, F. 1999. Cross-cultural adaptation of a health status classification system in children with cancer. First results of the French adaptation of the health utilities index marks 2 and 3. International *Journal of Cancer*, 12: 112-118.

157. Le Grand, J. 1987. Inequalities in health Some international comparisons. *European Economic Review*, 31: 182-191.

158. Leclerc, A. Fassin, D. Grandjean, H. Kaminski, M. & Lang, T. 2000. *Les inégalités sociales de santé*. La Découverte (Ed). INSERM. Série Recherches.

159. Lecluyse, A. 2007. Income-related health inequality in Belgium: a longitudinal perspective. The European *Journal of Health Economics*, 8(3): 237-243.

160. Lecluyse, A. & Cleemput, I. 2006. Making health continuous: implications of different methods on the measurement of inequality. *Health Economics*, 15(1): 99-104.

161. Lefranc, A., Pistolesi, N., & Trannoy, A. 2004. Le revenu selon l'origine sociale. *Economie Et Statistique*, 371: 49-82.

162. Leplège, A., Ecosse, A., Pouchot, J., Coste, J., & Perneger, T. V. 2001. Le questionnaire SF-36. *Manuel de l'utilisateur et guide d'interprétation des scores*. Paris.

163. Leplège, A., Ecosse, A., Verdier, A., & Perneger, T. V. 1998. The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. *Journal of Clinical Epidemiology*, 51(11): 1013-1023.

164. Lerman, R. I. & Yitzhaki, S. 1989. Improving the accuracy of estimates of Gini coefficients. *Journal of Econometrics*, 42(1): 43-47.

165. Leu, R. E. & Schellhorn, M. 2006. The evolution of income-related health inequalities in Switzerland over time. *CESif Economic Studies*.

166. Levy, H. & Paroush, J. 1974. Multi-period stochastic dominance. *Management Science*, 21(December).

167. Liang, J., Bennett, J. M., Whitelaw, N. A., & Maeda, D. 1991. The structure of self-reported physical health among the aged in the United States and Japan. *Medical Care*, 29(12): 1161-1180.

168. Lindeboom, M. & van Doorslaer, E. 2004. Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 23(6): 1083-1099.

169. Llena-Nozal, A. 2007. On the dynamics of health, work and socioeconomic status. Vrije Univertiteit Amsterdam.

170. Lollivier, S. 2001. Endogénéite d'une variable explicative dichotomique dans le cadre d'un modèle probit bivarié. *Annales D'Economie Et De Statistique*, 62: 251-269.

171. Lombrail, P. & Pascal, J. 2005. Inégalités sociales de santé et accès aux soins. *Sève*,(8): 31-39.

172. Lombrail, P., Pascal, J., & Lang, T. 2004. Accès au système de soins et inégalités sociales de santé : que sait-on de l'accès secondaire ? *Santé, Société Et Solidarité*, 2.

173. Long, M. J. & Marshall, B. S. 1999. The relationship between self-assessed health status, mortality, service use, and cost in a managed care setting. *Health Care Management Review*, 24(4): 20-27.

174. Long, S. J. 1997. *Regression models for categorical and limited dependent variables*. USA: Sage Publications.

175. Long, S. J. & Freese, J. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. College Station.

176. Mackenbach, J. P. 2006. Health inequalities: Europe in profile. Report EU, 272358 1p 2k Feb06.

177. Mackenbach, J. P. 2002. Income inequality and population health. *Britain Medicine Journal*, 324.

178. Mackenbach, J. P. 2005. Health and health care. In A. Börsch-Supan, A. Brugiavini, H. Jürges, J. P. Mackenbach, J. Siegrist, & G. Weber (Eds.), *Health, ageing and retirement in Europe - First results from the Survey of Health, Ageing and Retirement in Europe*: 81-149. Mannheim.

179. Mackenbach, J. P. & Kunst, A. E. 1997. Measuring the magnitude of socio economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Social Science and Medicine*, 44: 757-771.

180. Mackenbach, J. P., Kunst, A. E., Cavelaars, A. E. J. M., Groenhof, F., Geurts, J. J. M., & the EU Working Group on Socioeconomic Inequalities in Health 1997. Socioeconomic inequalities in morbidity and mortality in Western Europe. *The Lancet*, 349(June 7): 1655-1659.

181. Manderbacka, K., Lundberg, O., & Martikainen, P. 1999. Do risk factors and health behaviours contribute to self-ratings of health? *Social Science and Medicine*, 48: 1713-1720.

182. Marmot, M. & Wilkinson, R. 1999. *Social determinants of health*. Oxford.

183. Mathers, C. & Douglas, R. M. 1998. Measuring progress in population health and well-being. In R. Eckersley (Ed.), *Measuring progress: Is life getting better?*: 125-155. Collingwood Vic.

184. McDowell, I. 2006. *Measuring health - a guide to rating scales and questionnaires.* Oxford.

185. McLeroy, K., Bibeau, D., Steckler, A., & Glanz, K. 1988. An ecological perspective on health promotion programs. *Health Education Quarterly*, 15: 351-377.

186. Mechanic, D. 1986. The concept of illness behaviour: culture, situation and personal predisposition. *Psychological Medicine*, 16: 1-7.

187. Melchior, M., Berkman, L. F., Kawachi, I., Krieger, N., Zins, M., & Bonenfant, S. 2006a. Lifelong socioeconomic trajectory and premature mortality (35-65 years) in France: findings from the GAZEL Cohort Study. *Journal of Epidemiology and Community Health*, 60: 937-944.

188. Melchior, M., Lert, F., Martin, M., & Ville, I. 2006b. Socioeconomic position in childhood and in adulthood and functional limitations in midlife: data from a nationally-representative survey of French men and women. *Social Science and Medicine*, 63(11): 2813-2824.

189. Merleau-Ponty, M. 1945. *Phénoménologie de la perception.* Paris: N.R.F.

190. Merleau-Ponty, M. 1948. *Sens et non-sens.*

191. Mesrine, A. 2000. La surmortalité des chômeurs: un effet catalyseur du chômage ? *Economie Et Statistique*, 4: 33-48.

192. Mizrahi, A. & Mizrahi, A. 1985. Indicateurs de morbidité et facteurs sociodémographiques. Rapport IRDES, 657.

193. Mizrahi, A. & Mizrahi, A. 1997. Disparités sociales de morbidité en France. Rapport IRDES, 1177.

194. Moesgaard Iburg, K. M., Salomon, J. A., Tandon, A., & Murray, C. J. L. 2002. Cross-population comparability of physician-assessed and self-reported measures of health. In C. J. L. Murray, J. A. Salomon, C. Mathers, & A. D. Lopez (Eds.), *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*: 433-448. Geneva: WHO.

195. Monteil, C. & Robert-Bobée, I. 2005. Les différences sociales de mortalité: en augmentation chez les hommes, stables chez les femmes. *Insee Première*, INSEE, 1025.

196. Muller, C. & Trannoy, A. 2003. Multidimensional inequality comparisons: a compensation perspective. Unpublished Work, Mimeo THEMA(18).

197. Murray, C. J. L. & Chen, L. C. 1992. Understanding Morbidity Change. *Population and Development Review*, 18(3): 481-503.

198. Murray, C. J. L., Salomon, J. A., & Mathers, C. 2000. A Critical Examination of Summary Measures of Population Health. *Bulletin of the World Health Organisation*, 78(8): 981-994.

199. Nagyova, I. 2005. Self-rated health and quality of life in Slovak rheumatoid arthritis patients. Kosice Institute for Society and Health (KISH).

200. Newey, W. K. & West, K. D. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3): 703-708.

201. Noble, E. P., Zhang, X., Ritchie, T., Lawford, B. R., Grosser, S. C., Young, R. M., Sparkes, R. S., Keller, S. D., Ware, J. E., Gandek, B., Aaronson, N. K., Alonso, J., Apolone, G., Bjorner, J. B., Brazier, J., Bullinger, M., Fukuhara, S., Kaasa, S., Leplège, A., Sanson-Fisher, R. W., Sullivan, M., & Wood-Dauphinee, S. 1998. Testing the equivalence of translations of widely used response choice labels - The IQOLA Project approach. *Journal of Clinical Epidemiology*, 51(11): 933-944.

202. Nord, E. 1997. A review of synthetic health indicators. Reports OECD.

203. O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J. L., Dorian, P., & Brazier, J. 2003. A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics*, 12: 975-981.

204. OECD 2007. *Regards sur l'éducation 2007 - Les indicateurs de l'OCDE.* OCDE Publications.

205. Ok, E. A. & Kranich, L. 1998. The measurement of opportunity inequality: a cardinality-based approach. *Social Choice and Welfare*, 15: 263-287.

206. Ooghe, E. 2007. Sequential dominance and weighed utilitarianism. *Economics Letters*, 94: 208-212.

207. Parsons, D. O. 1980. The decline in male labor force participation. *Journal of Political Economy*, 88(1): 117-134.

208. Pigou, A. C. 1920. *The economics of welfare.* London: Macmillian and Co.

209. Poulton, R., Caspi, A., Milne, B. J., Thomson, M. W., Taylor, A., Sears, M. R., & Mofiett, T. E. 2002. Association between children's experience of socioeconomic disadvantage and adult health: a life-course study.*The Lancet*, 360: 1640-1645.

210. Power, C. & Hertzman, C. 1997. Social and biological pathways linking early life and adult disease. *British Medical Bulletin*, 53(1): 210-221.

211. Power, C., Matthews, S., & Manor, O. 1998. Inequalities in self-rated health: explanations from different stages of life. *The Lancet*, 351(4): 1009-1014.

212. Preston, S.H. Haines, M.R. & Pamuk, E. 1981. Effects of industrialization and urbanization on mortality in developped countries. Conference Proceeding of International Union for the Scientific Study of Population, Manila. 233-254.

213. Radloff, L. S. 1977. A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1: 385-401.

214. Rawls, J. 1982. Social utility and primary goods. In A. Sen & B. Williams (Eds.), *Utilitarianism and beyond*: 159-186. Cambridge: Cambridge University Press.

215. Raynaud, D. 2003. L'impact de la CMU sur la consommation individuelle de soins. *Etudes Et Résultats*, DREES, 229(mars).

216. Rochaix, L. & Hartmann, L. 2005. The French public-private mix. In A. K. Maynard (Ed.), *The public-private mix for health care.* Radcliffe.

217. Rodin, J. & MacAvay, G. 1992. Determinants of change in perceived health in a longitudinal study of older adults. *The Gerontologist*, 47(6): 373-384.

218. Roemer, J. 1998. *Equality of opportunity.* Harvard University Press. Cambridge.

219. Ruger, J. P. & Kim, H.-J. 2006. Global health inequalities: an international comparison. *Journal of Epidemiology and Community Health*, 60: 928-936.

220. Seibt, B. 1998. Kognitive Prozesse bei der Beurteilung der eigenen Gesundheit and die Validitat dieses Urteils. Free University of Berlin.

221. Sen, A. 1992. *Inequality Reexamined.* Oxford University Press

222. Sen, A. 1973. *On economic inequality.* Oxford University Press.

223. Sen, A. 1982. *Choice, welfare and measurement.* Basil Blackwell. Oxford.

224. Shmueli, A. 2003. Socio-economic and demographic variation in health and in its measures: the issue of reporting heterogeneity. *Social Science and Medicine*, 57(1): 125-134.

225. Shorrocks, A. 1983. Ranking income distributions. *Economica*, 50(197): 3-17.

226. Silventoinen, K., Lahelma, E., & Rahkonen, O. 2007. Social background, adult body-height and health. International *Journal of Epidemiology*, 28: 911-918.

227. Simon, J. G., van de, M. H., van der Meer, J. B., & Mackenbach, J. P. 2000. Socioeconomic differences in self-assessed health in a chronically ill population: the role of different health aspects. *J Behav.Med*, 23(5): 399-420.

228. Smith, J. P. 1999. Healthy bodies and thick wallets: the dual relation between health and economic status. *Journal of Economic Perspectives*, 13: 145-166.

229. Stewart, M. 1983. On least squares estimation when the dependent variable is grouped. *Review of Economic Studies*, L: 737-753.

230. Subramanian, S.V., & Kawachi, I. 2004. Income inequality and health: what have we learned so far? *Epidemiologic Reviews*, 26(1): 78-91.

231. Tandon, A. Murray, C.J.L. Salomon, J.A. & King, G. 2002. Statistical models for enhancing cross-population comparability. WHO. Global Programme for Evidence on Health Policy Discussion 42. Geneva.

232. The EuroQol Group 1990. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3): 199-208.

233. Thomas, D. & Frankenberg, E. 2002. The measurement and interpretation of health in social surveys. In C. J. L. Murray, J. A. Salomon, C. Mathers, & A. D. Lopez (Eds.), *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications.* Geneva: WHO.

234. Toulemon, L. 2003. La fécondité en France depuis 25 ans. Rapport pour le Haut Conseil de la Population et de la Famille.

235. Townsend, P. & Davidson, N. 1982. *Inequalities in health: the Black report*. Harmondsworth (UK).

236. Trannoy, A. 1999. Egalitarisme de la dominance et utilitarisme. *Revue Economique*, 50(4): 733-755.

237. Trannoy, A. 2006. Multidimensional egalitarism and the dominance approach: a lost paradise? In F. Farina & S. Savaglio (Eds.), *Inequality and Economic Integration*: 284-304.

238. Tubeuf, S. & Rochaix, L. 2007. Equité et santé: Apports de l'analyse économique et perspectives de recherche. Communication présentée aux JESF 2007 - Lille, France.

239. Vallin, J. & Meslé, F. 2001. Tables de mortalité françaises pour les XIXème et XXème siècle et projections pour le XXIème. Paris, INED. CD-ROM, *Données Statistiques*. [4].

240. van Doorslaer, E. & Gerdtham, U. G. 2003. Does inequality in self-assessed health predict inequality in survival by income?- Evidence from Swedish data. *Social Science and Medicine*, 57(9): 1621-1629.

241. van Doorslaer, E. & Jones, A. 2003. The determinants of self-reported health inequalities: validation of a new approach to measurement. *Journal of Health Economics*, 22: 61-87.

242. van Doorslaer, E. & Koolman, X. 2004. Explaining the differences in income-related health inequalities across European countries. *Health Economics*, 13(7): 609-628.

243. van Doorslaer, E., Koolman, X., & Jones, A. 2004. Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics*, 13(7): 629-647.

244. van Doorslaer, E., Koolman, X., & Puffer, F. 2002. Equity in the use of physicians visits in OECD countries: has equal treatment for equal need been achieved? Measuring Up: Improving Health Systems Performance in OECD Countries: 225-248.

245. van Doorslaer, E., Wagstaff, A., van de Burg, H., Christansen, T., De Graeve, D., Duchesne, I., Gerdtham, U. G., Gerfin, M., Geurts, J. J. M., Gross, L., Hakkinen, U., John, J., Klavus, J., Leu, R. E., Nolan, B., O'Donnell, O., Propper, C., Puffer, F., Schellhorn, M., Sundberg, G., & Winkelhake, O. 2000. Equity in the delivery of health care in Europe and the US. *Journal of Health Economics*, 19: 553-583.

246. van Ourti, T. 2003. Socioeconomic inequality in ill-health amongst the elderly. Should one use current or permanent income? *Journal of Health Economics*, 22: 219-241.

247. Veit, C. T. & Ware, J. E. 1983. The structure of psychological distress and well-being in general populations. *Journal of Consulting and Clinical Psychology*, 51: 730-742.

248. Verbrugge, L. M. 1984. Long life but worsening health? Trends in health and mortality of middle-aged and older persons. *Milbank Memorial Fund Quarterly/Health and Society*, 62(3): 475-519.

249. Villermé, L. R. 1830. De la mortalité dans les divers quartiers de la ville de Paris, et des causes qui la rendent très différentes dans plusieurs d'entre eux, ainsi que dans les divers quartiers de beaucoup de grandes villes. *Annales D'Hygiène Publique Et De Médecine Légale*, 3.

250. Wadsworth, M. E. J. 1999. Early life hypothesis. In M. Marmot & R. Wilkinson (Eds.), *Social determinants of health*: Oxford: Oxford University Press.

251. Wagstaff, A. 2005. The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics*, 14(4): 429-432.

252. Wagstaff, A. 2002. Inequality aversion, health inequalities and health achievement. *Journal of Health Economics*, 21: 627-641.

253. Wagstaff, A., Paci, P., & van Doorslaer, E. 1991. On the measurement of inequalities in health. *Social Science and Medicine*, 33(5): 545-557.

254. Wagstaff, A. & van Doorslaer, E. 1994. Measuring inequalities in health in the presence of multiple-category morbidity indicators. *Health Economics*, 3: 281-291.

255. Wagstaff, A. & van Doorslaer, E. 2000a. Equity in health care and delivery. In Culyer A.J & Newhouse J.P (Eds.), *Handbook of Health Economics* vol 1B: 1803-1857. Amsterdam: Elsevier.

256. Wagstaff, A. & van Doorslaer, E. 2000b. Income inequality and health: what does the literature tell us? *Annual Review of Public Health*, 21: 543-567.

257. Wagstaff, A. & van Doorslaer, E. 2003. Overall vs. socioeconomic health inequality: a measurement framework and two empirical illustrations. *Health Economics*, 13(3): 297-301.

258. Wagstaff, A., van Doorslaer, E., & Watanabe, N. 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics*, 112(1): 207-223.

259. Ware, J. E. 1986. The assessment of health status. In L. H. Aiken & D. Mechanic (Eds.), *Applications of Social Science to Clinical Medicine and Social Policy*.

260. Ware, J. E. & Sherbourne, C. D. 1992. The MOS 36-Item short-form health survey (SF-36): Conceptual framework and item selection. *Medical Care*, 30: 473-483.

261. Whitehead, M., Townsend, P., Davidson, N., & Davidsen, N. 1992. *Inequalities in health: The Black Report*. London.

262. Wilkinson, R. 1996. *Unhealthy societies: the afflictions of inequality*. New York.

263. Wilkinson, R. 1997. Commentary: income inequality summarises the health burden of individual relative deprivation. *British Medical Journal*, 314: 1727-1728.

264. World Bank. 2005. *World Development Report 2006: Equity and development*. Washington, D.C.

265. World Health Organisation. 2000. Rapport sur l'état de la santé dans le monde pour un système de santé plus performant. World Health Report. Geneva

266. World Health Organisation 1996. *Health interview surveys: towards international harmonization*. De Bruin, A. Picavet, H.S.V. & Nossikov, A. (Eds). WHO Regional Publications. 58

267. World Health Organisation 2003. *Social determinants of health. The solid facts*. Wilkinson,R. & Marmot,M. (Eds). Copenhagen.

268. World Health Organization 2002. *Summary measures of population health: concepts, ethics, measurement and applications*. Murray,C.J.L. Salomon, J.A. Mathers, C. & Lopez, A.D. (Eds). Geneva.

269. Yitzhaki, S. 2003. On an extension of the Gini inequality index. *International Economic Review*, 66(3): 541-562.

270. Zheng, B. 2006. Measuring health inequality and health opportunity. Unpublished Work, presented at the 2nd meeting of the Society for the Study of Economic Inequality (ECINEQ).