HAL
open science

# Les utilités EQ-5D et SF-6D dans la polyarthrite débutante

Cécile Gaujoux-Viala

## To cite this version:

Cécile Gaujoux-Viala. Les utilités EQ-5D et SF-6D dans la polyarthrite débutante. Médecine humaine et pathologie. Université de Lorraine, 2012. Français. NNT : 2012LORR0418 . tel-01751208

**HAL Id: tel-01751208**

**https://hal.univ-lorraine.fr/tel-01751208v1**

Submitted on 29 Mar 2018

**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 DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITÉ DE LORRAINE

## Ecole Doctorale BioSE (Biologie-Santé-Environnement)

## Thèse

**Présentée et soutenue publiquement pour l'obtention du titre de**

**DOCTEUR DE l'UNIVERSITE DE LORRAINE**

**Mention : « Sciences de la Vie et de la Santé »**

par **Cécile GAUJOUX-VIALA**

# Les utilités EQ-5D et SF-6D dans la polyarthrite débutante

**Le 7 novembre 2012**

**Membres du jury :**

| | | |
|---|---|---|
| **Rapporteurs :** | **Me Karine CHEVREUL** | **PhD, unite de cherche clinique en économie de la santé de l'AP-HP et au LIC EA 4393 université Paris 12, Paris** |
| | **Mr Jean-Pierre DAURES** | **PUPH, IURC, Laboratoire de Biostatistique d'Epidémiologie et de Recherche Clinique EA 2415, Montpellier** |
| **Examinateurs :** | **Mr Francis GUILLEMIN** | **PUPH, Lorraine-Université, Université Paris Descartes, EA 4360 Apemac, Nancy** |
| | **Mr Alain SARAUX** | **PUPH, EA 2216, Université de Bretagne occidentale, Brest** |
| | **Me Florence TUBACH** | **PUPH, Université Paris Diderot et INSERM CIE801, Paris** |
| | **Mr Bruno FAUTREL** | **PUPH, GRC-UPMC 08 – EEMOIS, Paris, directeur de thèse** |
| | **Me Anne-Christine RAT** | **MCUPH, Lorraine-Université, Université Paris Descartes, EA 4360 Apemac, Nancy, co-directeur de thèse** |

---

**EA 4360, Unité de recherche Maladies chroniques, santé perçue et processus d'adaptation. Approches épidémiologiques et psychologiques**

**UNIVERSITE HENRI POINCARE – NANCY I**

FACULTE DE MEDECINE

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE NANCY I

**Discipline : Epidémiologie et Santé publique**

présentée et soutenue publiquement

par

Cécile GAUJOUX-VIALA

Le 7 novembre 2012

TITRE

# Les utilités EQ-5D et SF-6D dans la polyarthrite débutante

DIRECTEURS DE THESE

Pr Bruno Fautrel

Dr Anne-Christine Rat

**Jury**

Professeur Karine CHEVREUL (Paris) …….…….Rapporteur

Professeur Jean-Pierre DAURES (Nîmes)……..…….Rapporteur

Professeur Francis GUILLEMIN (Nancy)….……..Examinateur

Professeur Alain SARAUX (Brest) ……...……..Examinateur

Professeur Florence TUBACH (Paris) ……Examinateur

Professeur Bruno FAUTREL (Paris) ……...Examinateur

Docteur Anne-Christine RAT (Nancy) ……Examinateur

# TABLE DES MATIERES

**REMERCIEMENTS** ..... **7**

**RESUME** ..... **9**

**ABSTRACT** ..... **11**

**LISTE DES PUBLICATIONS DE LA THESE** ..... **13**

**AUTRES PUBLICATIONS / COMMUNICATIONS DANS LE MEME DOMAINE** ... **15**

**PRESENTATION DES AXES DE RECHERCHE** ..... **19**

**SITUATION DES CONNAISSANCES** ..... **23**

**I – LA POLYARTHRITE RHUMATOÏDE (PR)** ..... **23**

I.1 – ASPECTS EPIDEMIOLOGIQUES DE LA PR ..... 23

I.2 – ASPECTS CLINIQUES DE LA PR ..... 23

I.3 – IMPACT DE LA PR ..... 27

I.3.1 – Douleur, incapacité et altération de la qualité de vie ..... 27

I.3.2 – Retentissement sur le travail ..... 28

I.3.3 – Augmentation de la mortalité ..... 28

I.4 – LES STRATEGIES THERAPEUTIQUES DANS LA PR ..... 30

I.4.1 – Les différents traitements de la PR ..... 30

I.4.2 – La notion de fenêtre thérapeutique ..... 31

I.5 – IMPACT ECONOMIQUE DE LA PR ..... 32

I.5.1- Contexte économique de la PR ..... 32

I.5.2- Evolution des coûts de la PR ces dernières années ..... 33

**II – L'EVALUATION MEDICO-ECONOMIQUE ..... 37**

II.1 – PRINCIPE ..... 37

II.2 – TYPOLOGIE DES COUTS ..... 39

II.2.1 – Coûts directs ..... 39

II.2.2 – Coûts indirects ..... 40

II.2.3 – Coûts intangibles ..... 40

II.3 – TYPOLOGIE DES GAINS DE SANTE ..... 40

II.3.1 – Une vision comptable : gain exprimé en unité monétaire ..... 41

II.3.2 – La vision des médecins: gain d'efficacité exprimé en unité naturelle ..... 41

II.3.3 – Gain d'utilité: un nécessaire compromis ..... 42

**III – LA NOTION D'UTILITE ..... 43**

III.1 – PRINCIPES ..... 44

III.1.1 – Les préférences individuelles en santé ..... 44

III.1.2 – Le QALY ..... 45

III.2 – METHODES D'OBTENTION DE L'UTILITE ..... 46

III.2.1 – Utilité directe ..... 47

III.2.1.1 – Jeu de loterie ..... 47

III.2.2.2 – Arbitrage temporel ..... 48

III.2.2 – Utilité indirecte ..... 49

III.2.2.1 – EQ-5D ..... 50

III.2.2.2 – SF-6D ..... 50

III.3 – AVANTAGES DE L'UTILITE ET DES QALYS ..... 52

III.4 – LIMITES DE L'UTILITE ET DES QALYS ..... 52

**RESULTATS DES TRAVAUX DE RECHERCHE ..... 57**

**I – PROPRIETES PSYCHOMETRIQUES DE L'EQ-5D ET DU SF-6D EN TRANSVERSAL..... 59**

**II – SENSIBILITE AU CHANGEMENT DE L'EQ-5D ET DU SF-6D ..... 63**

**III – LES PATIENTS AVEC UNE UTILITE NEGATIVE SUR L'EQ-5D, UN ETAT DE SANTE « PIRE QUE LA MORT ..... 67**

**IV – DETERMINANTS DE L'UTILITE EQ-5D, SF-6D ET DE LA DIFFERENCE ENTRE LES DEUX . 71**

**V – EVOLUTION DES DETERMINANTS DE L'UTILITE EQ-5D, SF-6D ET DE LA DIFFERENCE ENTRE LES DEUX AU COURS DU TEMPS ..... 77**

**VI – LES CHANGEMENTS D'UTILITE AVEC L'EQ-5D ET LE SF-6D ..... 81**

**SYNTHESE..... 89**

**PERSPECTIVES..... 97**

**I – DANS LA POLYARTHRITE RHUMATOÏDE : ..... 97**

**I I– AUTRES MALADIES RHUMATOLOGIQUES : ..... 99**

II.1 – SPONDYLARTHRITE ANKYLOSANTE..... 99

II.2 – L'ARTHROSE ..... 99

**REFERENCES ..... 101**

**ANNEXES..... 109**

# REMERCIEMENTS

**Au Professeur Bruno Fautrel**,
Qui m'a donné envie de faire de la Rhumatologie et avec qui je travaille avec bonheur depuis trois ans. Jeune interne, je l'ai connu « intermittent du spectacle » comme il s'était lui-même présenté, il est maintenant professeur et chef de service et il a su garder sa modestie et son enthousiasme.
Qu'il trouve ici l'expression de ma profonde et très sincère admiration et amitié.

**Au Docteur Anne-Christine Rat,**
Qui a su m'accompagner, me conseiller, et m'encourager tout au long de ce travail de recherche. Je souhaite tout particulièrement t'exprimer ma reconnaissance pour m'avoir fait bénéficier de ton expérience et m'avoir accordé ta confiance.

**Aux Professeur Francis Guillemin, Serge Briançon et toute l'équipe de l'Ecole de Santé Publique de Nancy**,
Qui m'ont accueillie pendant ces 3 années et ont su grâce à leurs remarques pertinentes me faire progresser.

**Aux Professeurs Karine Chevreuil et Jean-Pierre Daurès**,
Qui m'ont fait l'honneur d'accepter d'être les rapporteurs de ce travail de thèse.

**Aux Professeurs Florence Tubach et Alain Saraux**,
Qui m'ont fait l'honneur d'accepter d'en être les examinateurs.

**Au Professeur Pierre Bourgeois et à toute l'équipe de Rhumatologie de la Pitié Salpêtrière**,
Avec qui j'ai eu le plaisir de travailler durant ces 3 années.

**Au Professeur Jean-Louis Leroux et à toute l'équipe de Rhumatologie du CHU de Nîmes,**
Qui vont très bientôt m'accueillir dans leur service.

**Aux Professeurs Maxime Dougados, Bernard Combe, Jacques Morel, Xavier Chevalier et Jean Sibilla**,
Qui m'ont soutenue et encouragée dans mes projets de carrière universitaire.

**Au Docteur Laure Gossec**, dont le dynamisme et l'enthousiasme me servent d'exemple, arrivant à concilier vie de famille et carrière universitaire.

**A mes parents.** Merci pour votre amour inconditionnel et votre indéfectible soutien.

**A ma grand-mère** que j'aime « beaucoup et tendrement ».

**A mon grand-père** qui m'a appris « d'abord, on travaille, ensuite, on peut s'amuser » mais qui disait aussi qu' « il n'y a pas beaucoup d'occasions de rire et quand on en a une, il ne faut pas la rater ». J'espère que de là où tu es, tu peux voir le résultat de ces années de travail.

**A mon frère Pierre et ma sœur Marie** qui partagent mon amour de la médecine.

**A ma famille,** tendrement.

**A mes amis,** Bénédicte, Mélanie, Amélie, Jonathan, Julie, Nico et Benjamin
Qui malgré l'éloignement sont restés présents dans ma vie.

**Aux deux hommes de ma vie, Thomas et Alexandre**.

**Je dédie cette thèse à Alexandre**, mon petit amour,
Qui est l'exact reflet temporel de ce travail et qui a toujours tout fait pour faciliter la vie de sa maman.

## RESUME

La notion d'instauration rapide d'un traitement efficace et agressif dans la polyarthrite débutante a renforcé la nécessité de disposer de mesures d'utilité validées afin d'optimiser l'évaluation médico-économique de ces traitements. Actuellement, il n'y a pas de consensus sur le choix du questionnaire multi-attributs à utiliser pour le calcul des utilités. L'objectif de cette thèse est donc d'étudier les niveaux, les performances et les déterminants des 2 outils indirects de mesure d'utilité (EQ-5D et SF-6D) dans la polyarthrite débutante, au travers des données de la cohorte ESPOIR, cohorte multicentrique française de polyarthrites débutantes.

La concordance entre les deux instruments d'utilité est faible chez les patients aux extrémités de l'échelle: en cas de bonne santé, l'utilité EQ-5D est systématiquement supérieure à l'utilité SF-6D, par contre, si les patients vont mal, elle est toujours inférieure. La validité de construction et la capacité discriminante sont bonnes et quasi-similaires entre les deux instruments. Par contre, la sensibilité au changement diffère selon le sens du changement : le SF-6D est plus sensible à l'amélioration et l'EQ-5D est plus sensible à la détérioration. La moyenne de changement en valeur absolue est plus importante pour l'EQ-5D que pour le SF-6D. Le gain de QALYs pour une même stratégie thérapeutique est doublé avec l'EQ-5D. Le SF-6D est plus conservatoire. Enfin, l'estimation du gain d'utilité avec le SF-6D est plus précise car sa variabilité est toujours plus faible qu'avec l'EQ-5D. Sur ces premiers arguments, l'utilisation du SF-6D paraît plus appropriée pour l'évaluation médico-économique des nouveaux traitements chez les patients souffrant d'arthrite débutante.

L'EQ-5D et le SF-6D donnent des résultats différents. Notamment l'EQ-5D autorise des utilités négatives, états de santé « pire que la mort », déterminés par la fonction (HAQ) et l'état mental en analyse multivariée et où la gêne/douleur est maximale. La fonction et l'état mental sont en fait les déterminants principaux de l'utilité EQ-5D et SF-6D. Une incapacité fonctionnelle importante et un mauvais score mental sont les éléments clés conduisant à une différence entre les deux mesures d'utilité. La variabilité des déterminants du SF-6D, de l'EQ-5D et des différences entre les 2 instruments d'utilité au cours du temps, les difficultés de modélisation linéaire de l'EQ-5D posent question quant à l'utilisation du mapping pour estimer l'utilité à partir d'instruments de mesure de l'état de santé notamment à partir du HAQ.

Mots clés : polyarthrite rhumatoïde – utilité indirecte – EQ-5D – SF-6D – évaluation.

## ABSTRACT

The explosion of drug development for rheumatoid arthritis and the revolution of early aggressive therapy for the disease have fuelled the search for better approaches to establish cost-effectiveness in early arthritis (EA), but consensus is lacking on the choice of utility instrument. We aimed to compare the EQ-5D and SF-6D, 2 indirect utility measures widely used to calculate quality-adjusted life-years (QALYs), in terms of their utility values, their performance and their determinants in a large prospective cohort of patients with EA.

Although the correlation between the 2 utility scores was high, descriptive statistics revealed systematic disagreement at both the low and high ends of the utility scales: the utility obtained was systematically higher with the SF-6D than the EQ-5D for patients with a low health state. In contrast, for patients with a high health state, the EQ-5D score was systematically higher than that of the SF-6D. Their construct validity and their discrimination capacity were good and similar. The responsiveness of the EQ-5D and SF-6D differed according to the direction of change in EA condition: the SF-6D was more responsive to improved condition and the EQ-5D to deteriorated condition. The high mean change of the EQ-5D relative to the SF-6D has consequences for cost-effectiveness analyses: the QALYs gain demonstrated by the EQ-5D was near twice that of the SF-6D. The SF-6D was more conservative than the EQ-5D. The smaller variance of the SF-6D should result in less uncertainty in estimating the relative cost-effectiveness of 2 treatments. Therefore, the SF-6D may be more appropriate for use in randomized controlled trials of biologic treatments for patients with EA.

EQ-5D and SF-6D are not interchangeable. The EQ-5D allows for negative utility values, which theoretically correspond to health states “worse than death" (not available with SF-6D), with extreme pain/discomfort. Poor HAQ and SF-36 mental component scores were always independently associated with negative EQ-5D. The function and the mental state are in fact the main determinants of SF-6D and EQ-5D utility measures. Higher functional disability and lower mental status were the key element leading to substantial difference between the two utility measures in EA patients. The variability of the determinants of SF-6D and EQ-5D utility measure and of the difference between SF-6D and EQ-5D over 3 years in addition of the difficulty of modeling EQ-5D by linear regression raise concerns about mapping to estimate utilities from clinical instruments.

Key words: rheumatoid arthritis – indirect utility – EQ-5D – SF-6D – evaluation.

## LISTE DES PUBLICATIONS DE LA THESE

1 – Comparison of the EQ-5D and the Short Form 6D utility measures in French patients with early arthritis: Results from the ESPOIR cohort
**Gaujoux-Viala C**, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, Fautrel B
J Rheumatol. 2011 Aug;38(8):1576-84.

2 – Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort.
**Gaujoux-Viala C**, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, Fautrel B.
Ann Rheum Dis. 2012 Sep;71(9):1478-83.

3 – Who are the patients with early arthritis with "worse than death" scores on the EQ-5D? Results from the ESPOIR cohort
**Gaujoux-Viala C**, Fautrel B, Guillemin F, Flipo RM, Bourgeois P, Rat AC
Rheumatology 2012 [Epub ahead of print]

4 – The impact of functional disability on EQ-5D and SF-6D utility discrepancies in early arthritis: Results from the ESPOIR cohort
**Gaujoux-Viala C**, Fautrel B, Hosseini K, Guillemin F, Flipo RM, Rat AC.
Soumis

5 – The determinants of the SF-6D and EQ-5D utility measures vary over time in early arthritis:Results from the ESPOIR cohort
**Gaujoux-Viala C**, Rat AC, Hosseini K, Guillemin F, Flipo RM, Fautrel B.
Article en cours de soumission

6 – Change in two health state utility measures, EQ-5D and SF-6D, and comparison of their QALYs estimates in early arthritis: Results from the ESPOIR cohort
**Gaujoux-Viala C**, Fautrel B, Hosseini K, Guillemin F, Flipo RM, Rat AC.
Manuscript en cours de préparation

## AUTRES PUBLICATIONS / COMMUNICATIONS DANS LE MEME DOMAINE

### *Articles dans des journaux avec comité de lecture*

1. Hosseini K, **Gaujoux-Viala C**, Guillemin Francis. Measuring change of utility within patient: relevance and integration in cost-utility analysis. Soumis à Quality of life research
2. Hosseini K, **Gaujoux-Viala C**, Coste J, Fautrel B, Rat Anne-Christine,Guillemin Francis The impact of comorbidities on the measurement of indirect utility by SF-6D in lower limb osteoarthritis. Best Practice & Research Clinical Rheumatology. In press
3. **Gaujoux-Viala C,** Fautrel B. Cost-effectiveness of therapeutic interventions in ankylosing spondylitis: A critical and systematic review. Accepté dans Pharmacoeconomics.

### *Recommandations*

1. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, **Gaujoux-Viala C**, Gorter S, Knevel R, Nam J, Schoels M, Aletaha D, Buch M, Gossec L, Huizinga T, Bijlsma JW, Burmester G, Combe B, Cutolo M, Gabay C, Gomez-Reino J, Kouloumas M, Kvien TK, Martin-Mola E, McInnes I, Pavelka K, van Riel P, Scholte M, Scott DL, Sokka T, Valesini G, van Vollenhoven R, Winthrop KL, Wong J, Zink A, van der Heijde D.
   EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs.
   Ann Rheum Dis. 2010 Jun;69(6):964-75.

2. **Gaujoux-Viala C**, Smolen JS, Landewé R, Dougados M, Kvien TK, Mola EM, Scholte-Voshaar M, van Riel P, Gossec L. Current evidence for the management of rheumatoid arthritis with synthetic disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. Ann Rheum Dis. 2010 Jun;69(6):1004-9.

### *Ouvrage didactique*

1. **Gaujoux-Viala C,** Fautrel B**.** Evaluation médico-économique des biothérapies. DIU biothérapies du CRI 2012

***Communications orales (en 1er auteur) lors de congrès***

**2011**

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel
Qui sont les patients souffrant de polyarthrite débutante avec un état de santé "pire que la mort" sur l'EQ-5D ? Résultats de la cohorte ESPOIR. Congrès Français de Rhumatologie, Paris 2011

- **C Gaujoux-Viala**
Epidemiology is fun ! Next generation.
Oratrice invitée. Congrès Européen de Rhumatologie EULAR, Londres 2011

**2010**

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel
La comparaison de la sensibilité au changement de l'EQ-5D et du SF-6D chez les patients souffrant d'arthrite débutante est en faveur de l'utilisation du SF-6D: Résultats de la cohorte ESPOIR. Congrès Français de Rhumatologie, Paris 2010

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel
Comparaison de l'EQ-5D et du SF-6D chez les patients souffrant de polyarthrite débutante: Résultats de la cohorte ESPOIR. 4ème Conférence d'Epidémiologie Clinique, Paris 2010.

- **C Gaujoux-Viala**, F Gandjbakhch, B Fautrel
Faut-il utiliser les biothérapies en première intention dans la PR ?
55èmes Journées du Centre Viggo Petersen, Paris 2010

**2009**

- **C Gaujoux-Viala**, L Gossec
Synthetic DMARDS including combination DMARDS without glucocorticoid
Congrès Européen de Rhumatologie EULAR, Copenhague 2009

***Communications affichées (en 1er auteur) lors de congrès***

**2012**

- **C Gaujoux-Viala**, B Fautrel, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, AC Rat
Variability of the SF-6D determinants over time in early arthritis: results from the Espoir cohort. Congrès Américain de Rhumatologie ACR, Washington 2012

- **C Gaujoux-Viala**, B Fautrel, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, AC Rat
Substantial functional disability is the key determinant of discrepancies between EQ-5D and SF-6D utility measures in early arthritis: results from the Espoir cohort. Congrès Américain de Rhumatologie ACR, Washington 2012

- **C Gaujoux-Viala**, B Fautrel, K Hosseini, RM Flipo, F Guillemin, AC Rat
Une plus grande incapacité fonctionnelle conduit à une utilité plus faible par l'EQ-5D que par le SF-6D dans la polyarthrite débutante: résultats de la cohorte ESPOIR. 6ème Conférence Francophone d'Epidémiologie Clinique, Lyon 2012

- **C Gaujoux-Viala**, B Fautrel , F Guillemin, RM Flipo, P Bourgeois, AC Rat
Qui sont les patients souffrant de polyarthrite débutante avec un état de santé "pire que la mort" sur l'EQ-5D? Résultats de la cohorte ESPOIR. 6ème Conférence Francophone d'Epidémiologie Clinique, Lyon 2012

**2011**

- **C Gaujoux-Viala**, B Fautrel, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, AC Rat
Who are the patients with early arthritis who score states "worse than death" on the EQ-5D? Results from the ESPOIR cohort. Congrès Américain de Rhumatologie ACR, Chicago 2011

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel
La comparaison de la sensibilité au changement de l'EQ-5D et du SF-6D chez les patients souffrant d'arthrite débutante est en faveur de l'utilisation du SF-6D: Résultats de la cohorte ESPOIR. 5ème Conférence Francophone d'Epidémiologie Clinique, Marseille 2011. Récompensée par le *Prix jeune chercheur*.

- **C Gaujoux-Viala**, B Fautrel

Cost-effectiveness of therapeutic interventions in ankylosing spondylitis: A critical and systematic review. Congrès Européen de Rhumatologie EULAR, Londres 2011

**2010**

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel

The comparative responsiveness of the EQ-5D and SF-6D to change in patients with early Arthritis: Results from the ESPOIR cohort

Congrès Américain de Rhumatologie ACR, Philadelphie 2010

- **C Gaujoux-Viala**, AC Rat, F Guillemin, RM Flipo, P Fardellone, P Bourgeois, B Fautrel

How do the EQ-5D and the short form 6D utility measures compare in patients with early arthritis: results from the ESPOIR cohort

Congrès Européen de Rhumatologie EULAR, Rome 2010

## PRESENTATION DES AXES DE RECHERCHE

Depuis quelques années, en complément des traditionnels indices de morbidité et de mortalité, de nouveaux instruments de mesure de l'état de santé ont été développés. Si l'évaluation des interventions en santé est restée très longtemps fondée sur la mise en balance du bénéfice purement clinique pour les malades au regard des effets indésirables potentiels, la mesure plus globale des états de santé, intégrant la qualité de vie liée à la santé est devenue essentielle pour quantifier un gain de santé. Ceci est d'autant plus vrai dans le contexte actuel où des thérapeutiques de plus en plus coûteuses se heurtent à des ressources financières de plus en plus limitées, ce qui nécessite de faire des choix si possible éclairés (notion de coût d'opportunité ou de renonciation).

L'évaluation des états de santé doit concilier différents aspects :

- la vision du patient, expert de la maladie
- la vision de la société (population générale), pourvoyeur de ressources
- une évaluation chiffrée nécessaire pour la modélisation
- une unité commune pour toutes les maladies

L'utilité sert à évaluer la valeur relative de chaque état de santé. Elle permet d'attribuer une valeur numérique à un état de santé, variant de 0, à priori la mort, à 1, état de parfaite santé. Elle permet de calculer des espérances de vie pondérées par la qualité de vie (QALYs) indispensables pour rapprocher coûts et résultats dans les études coût-utilité.

L'évaluation des utilités liées à un état de santé donné peut être faite en recherche clinique grâce à des questionnaires multi-attributs tels que L'EQ-5Det le SF-6D, faciles à administrer et supposés équivalents en termes de propriétés métrologiques. Ils mesurent la perception des patients, mais aussi celle de la société à travers les poids sociétaux (pondérations issues d'enquêtes en population) qui sont affectés à divers états décrits par les individus. Bien que le concept théorique d'utilité sous-tende qu'à chaque état de santé corresponde une seule valeur d'utilité indépendamment de la façon de la mesurer, différents instruments peuvent donner différentes valeurs. Il existe très peu de données dans la polyarthrite débutante et actuellement, il n'y a pas de consensus sur le choix du questionnaire multi-attributs à utiliser pour le calcul des utilités. Or ce choix peut clairement modifier les résultats des études sur les nouveaux agents biologiques, les différentes stratégies thérapeutiques et leur rapport coût-utilité.

C'est l'objectif de cette thèse d'université : étudier les niveaux, les performances et les déterminants des 2 outils indirects de mesure d'utilité (EQ-5D et SF-6D) dans la polyarthrite débutante afin de déterminer si un des deux instruments est plus pertinent.

Dans la première partie consacrée à la contextualisation du problème, le premier chapitre sera consacré à la polyarthrite rhumatoïde (PR), maladie qui a servi de support à ce travail de recherche, notamment sur:

- ses aspects épidémiologiques et cliniques, nécessaires à une bonne compréhension de l'impact de la maladie, tant à l'échelon individuel qu'à celui de la société ;
- les traitements actuellement disponibles et les progrès thérapeutiques des dernières années.
- La notion de fenêtre thérapeutique avec l'idée actuelle de traiter « vite et fort »
- son impact économique

Le deuxième chapitre positionnera mes travaux de recherche dans le champ général de l'évaluation médico-économique avec :

- Un rappel sur les principes sous-tendant toute évaluation médico-économique ;
- Une présentation rapide des méthodes permettant de mesurer les coûts, c'est-à-dire les ressources mobilisées (*inputs*);
- Une description des différents moyens d'évaluer les bénéfices, les gains de santé (*outputs*), avec pour objectif d'établir des préférences sur lesquelles fonder des décisions de santé.

Le troisième chapitre développera la notion d'utilité, qui est un des moyens d'exprimer un gain de santé. Seront successivement abordés :

- Les principes généraux sous-tendant le concept d'utilité et les QALYs ;
- Les modalités d'obtention de l'utilité, avec notamment l'utilité directe (jeu de loterie, arbitrage temporel) et l'utilité indirecte (EQ-5D et SF-6D) ;
- La présentation en détails des 2 outils d'utilité indirecte utilisés dans le cadre de ce travail, l'EQ-5D et le SF-6D
- Les avantages et les limites de l'utilité et des QALYs;

Dans la deuxième partie, les travaux de recherche constitutifs de la thèse seront exposés, correspondant soit à des articles déjà rédigés, soit à la présentation de résultats faisant l'objet d'un manuscrit en cours de préparation. Tous les travaux ont été conduits sur les données de la cohorte ESPOIR, une cohorte multicentrique française de polyarthrites débutantes.

- Le premier travail a porté sur l'étude des propriétés psychométriques de l'EQ-5D et du SF-6D dans une approche transversale. Ont été évaluées la concordance entre les deux scores (coefficients de corrélations intra-classe [ICC], graphes de Bland-Altman), la validité de construction/construit (corrélation avec différents marqueurs de l'état de santé d'un patient souffrant de polyarthrite) et la capacité discriminante (efficacité relative).

- Le deuxième travail abordait l'étude longitudinale de l'EQ-5D et du SF-6D dans la cohorte ESPOIR. L'objectif de ce travail était de décrire et de comparer la sensibilité au changement de l'EQ-5D et du SF-6D dans une large cohorte multicentrique de patients souffrant de polyarthrite débutante.

- L'EQ-5D et le SF-6D donnent des niveaux d'utilité différents. Notamment l'EQ-5D autorise des utilités négatives donc des états de santé « pires que la mort » ce que ne permet pas le SF-6D. L'objectif de cette 3° étude était d'identifier les patients avec EQ-5D négatif dans la cohorte de polyarthrites débutantes ESPOIR et de décrire leurs caractéristiques. Le but était de déterminer quels aspects spécifiques (socio-démographiques, activité, incapacité, psychologique…) de la polyarthrite débutante étaient associés aux états de santé « pires que la mort ».

- La 4° étude portait sur le sens de l'utilité et de la différence d'utilité entre EQ-5D et SF-6D. Les objectifs de cette étude transversale était de 1) déterminer quelles variables étaient associées avec l'utilité EQ-5D et SF-6D dans la polyarthrite débutante 2) déterminer quelles variables étaient associées à une différence substantielle d'utilité entre l'EQ-5D et le SF-6D à l'inclusion dans la cohorte.

- Dans la 5° étude, nous nous sommes ensuite intéressés à l'évolution au cours du temps des déterminants de l'utilité EQ-5D, SF-6D et de la différence d'utilité entre les deux mesures.

- Enfin, le 6° et dernier travail longitudinal a porté sur la finalité des utilités, au travers de l'étude des changements d'utilité EQ-5D et SF-6D au cours du temps. Les gains de QALYs avec l'EQ-5D et le SF-6D dans l'ensemble de la population et sous différentes stratégies thérapeutiques ont également été calculés et comparés.

Cette thèse se terminera par une synthèse de ces principaux résultats, de leur conséquence et l'abord des perspectives et développements futurs de ces travaux.

# SITUATION DES CONNAISSANCES

## I – LA POLYARTHRITE RHUMATOÏDE (PR)

### I.1 – Aspects épidémiologiques de la PR

La PR est le plus fréquent des rhumatismes inflammatoires de l'adulte. Sa prévalence est estimée entre 0,3 et 1% de la population générale adulte, les femmes étant presque deux fois plus touchées que les hommes (1). Elle apparaît généralement entre 35 et 55 ans. Une méta-analyse de 2006 sur l'épidémiologie de la PR a identifié 28 études : 9 études d'incidence, 17 études de prévalence et 2 études estimant à la fois l'incidence et la prévalence (2). L'incidence annuelle médiane était de 16,5 cas pour $10^5$ sujets (extrêmes 9 à 24) pour les pays d'Europe du Sud, de 29 (extrêmes 24 à 36) pour les pays d'Europe du Nord et de 38 (extrêmes 31 à 45) pour les Etats-Unis. La prévalence médiane était de 0,33 % (extrêmes 0,31 à 0,5) pour les pays d'Europe du Sud, de 0,5 (extrêmes 0,44 à 0,8) pour les pays d'Europe du Nord et de 0,35 (extrêmes 0,24 à 0,36) pour les pays en voie de développement. Une étude conduite aux Etats-Unis a retrouvé une prévalence estimée à 1,07 %. Il semble exister une tendance à une moindre prévalence et incidence de la maladie au cours de la seconde partie du XXème siècle (2). Mais ces données reposent sur un faible nombre d'études, utilisant une méthodologie de détection des cas et des marqueurs ou critères diagnostiques différents.

### I.2 – Aspects cliniques de la PR

La PR est un rhumatisme qui se développe classiquement de manière bilatérale et symétrique en touchant préférentiellement les petites articulations des pieds et des mains, ainsi que les poignets. C'est une maladie chronique dont l'évolution se caractérise par des poussées dont la fréquence et l'intensité varient d'un patient à l'autre. La PR se manifeste par des douleurs et des gonflements des articulations. Les douleurs peuvent être intenses et sont principalement présentes la nuit et le matin au réveil. Les gonflements correspondent à la présence d'une atteinte inflammatoire de la synoviale (synovite) (Figure 1).

**Figure 1 :** Exemples de gonflements articulaires liés à la PR (synovites)

La PR est une maladie auto-immune systémique évolutive caractérisée par une inflammation chronique de la membrane synoviale des articulations, pouvant mener à des déformations (Figure 2) et une destruction articulaire (Figure 3).

**Figure 2 :** Exemples de déformations articulaires liées à la PR

**Figure 3 :** Destructions articulaires liées à la PR (mains à gauche, pieds à droite)

Il a été montré que ces dégâts articulaires surviennent précocement au cours de la PR : 70% des patients ont des érosions dès les 2 premières années de la maladie (3).

Le diagnostic de PR repose sur un faisceau d'arguments cliniques et biologiques. Les critères ont été récemment actualisés (Figure 4) (4). Le diagnostic de PR est d' autant plus probable qu'il existe un grand nombre d'articulations atteintes, un syndrome inflammatoire biologique, une durée d'évolution supérieure à 6 semaines et une positivité du bilan immunologique : facteur rhumatoïde (FR) et anticorps antipeptides citrullinés (anti CCP) (Tableau 1).

**Figure 4 :** Critères de l'ACR/EULAR 2010 pour le diagnostic de PR (4)

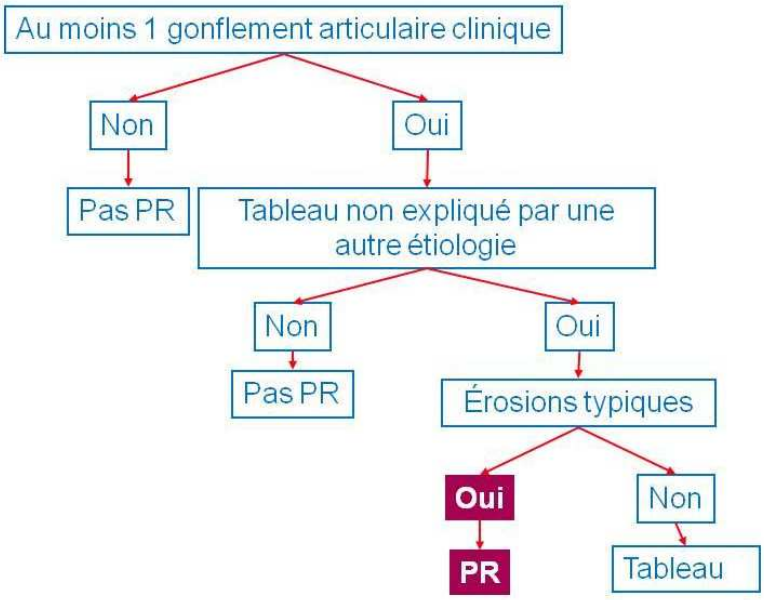



**Tableau 1 :** Critères de l'ACR/EULAR 2010 pour le diagnostic de PR en l'absence d'érosion typique (4)

| | |
|---|---|
| **A: Nombre d'articulations atteintes** | |
| - 1 grosse articulation. | 0 |
| - 2 à 10 grosses articulations. | 1 |
| - 1à 3 petites articulations. | 2 |
| - 4 à 10 petites articulations. | 3 |
| - > de 10 articulations (avec au moins une petite). | 5 |
| **B: Sérologie** | |
| - FR et ACPA négatifs. | 0 |
| - FR ou ACPA positif à un faible taux. | 2 |
| - FR ou ACPA positif à un taux élevé. | 3 |
| **C: Paramètres de l'inflammation** | |
| - VS et CRP normales. | 0 |
| - Elévation de la VS ou de la CRP. | 1 |
| **D: Durée d'évolution des symptômes** | |
| - Moins de 6 semaines. | 0 |
| - Plus de 6 semaines. | 1 |
| PR si score > ou = à 6 | |

## I.3 – Impact de la PR

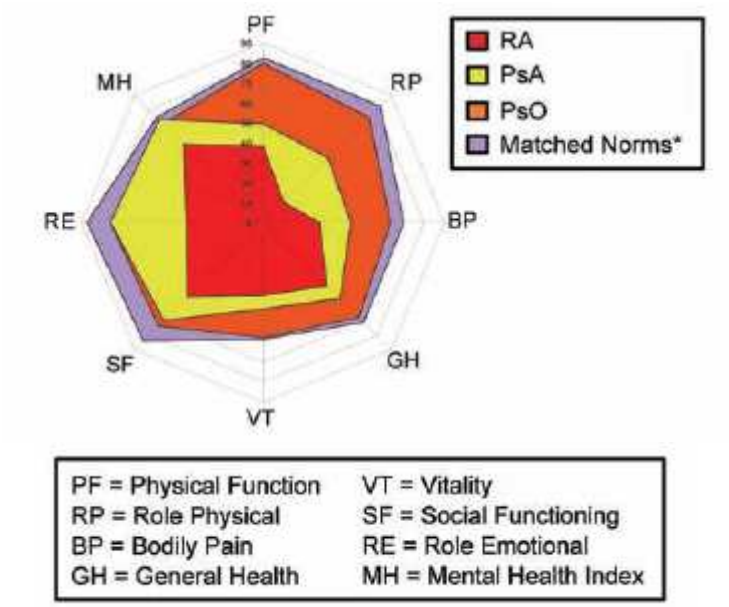
### I.3.1 – *Douleur, incapacité et altération de la qualité de vie*

Du fait des douleurs et des déformations articulaires, la PR est à l'origine d'un handicap croissant. Une étude menée en Hollande sur 186 PR a montré que, dès les premières années, la maladie a un retentissement dans la vie quotidienne chez 89% des patients. A trois ans d'évolution, 52% des patients avaient des difficultés dans leur déplacement (5).

La qualité de vie est significativement altérée dans la PR : douleur, fatigue, incapacité fonctionnelle souvent responsable de handicap (6).

Une étude a comparé la qualité de vie dans la PR (n=528), le rhumatisme psoriasique (n=583), le psoriasis (n=205) avec celle de sujets de la population générale américaine appariés sur l'âge et le sexe (7). Les patients étaient issus d'essais contrôlés randomisés. Toutes les dimensions de la qualité de vie du SF36 étaient plus altérées dans la PR (Figure 5).

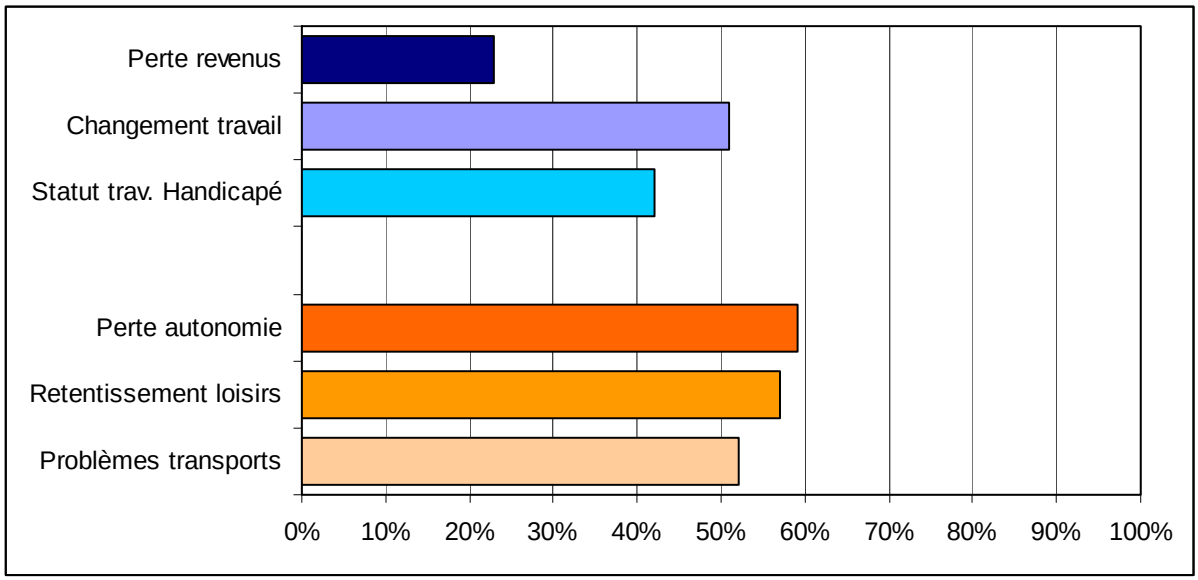
**Figure 5 :** Impact de la PR, du rhumatisme psoriasis et du psoriasis sur la qualité de vie (SF36) (7)



### I.3.2 – *Retentissement sur le travail*

L'impact de la PR sur le travail est important et ce dès le début de la maladie. Une étude menée en Hollande sur 186 PR a montré que 51% des patients présentaient une altération de leur productivité au travail dès la 1° année avec 25% des patients reconnus comme travailleurs handicapés (5). A trois ans d'évolution, 42% étaient reconnus comme travailleurs handicapés et 25% avaient une baisse de leur revenu (Figure 6). Dans la cohorte française de polyarthrites débutantes ESPOIR, 20% des patients étaient en congé maladie et 12% en invalidité 3 ans après le début de la maladie (8). Au bout de 8 ans, moins de 50% des patients atteints de polyarthrite peuvent continuer à travailler ou à vaquer normalement à leurs activités quotidiennes (9).

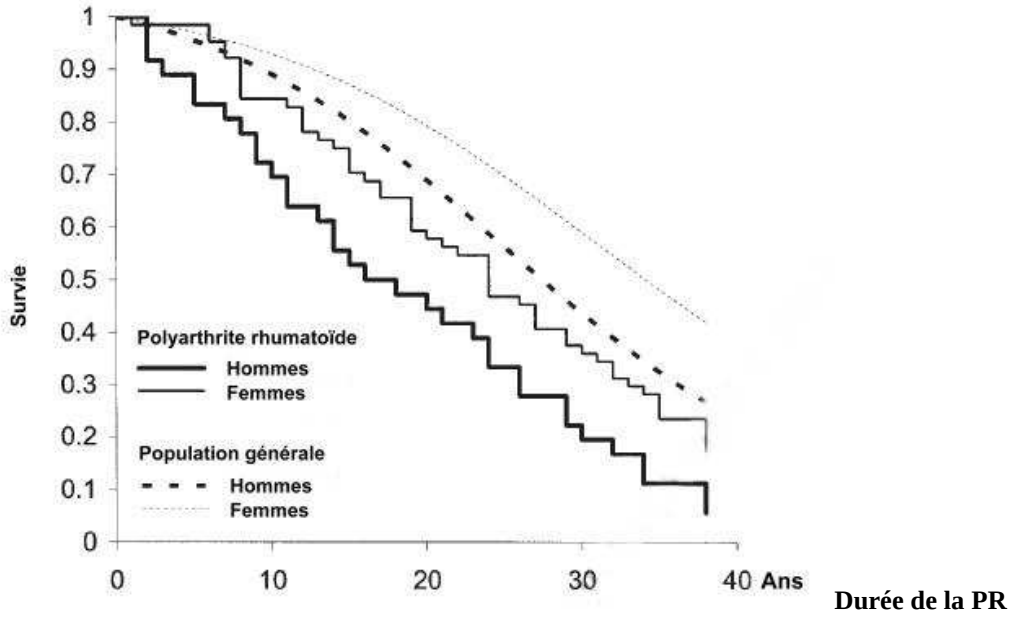
**Figure 6 :** Impact social de la PR d'après (5)




### I.3.3 – *Augmentation de la mortalité*

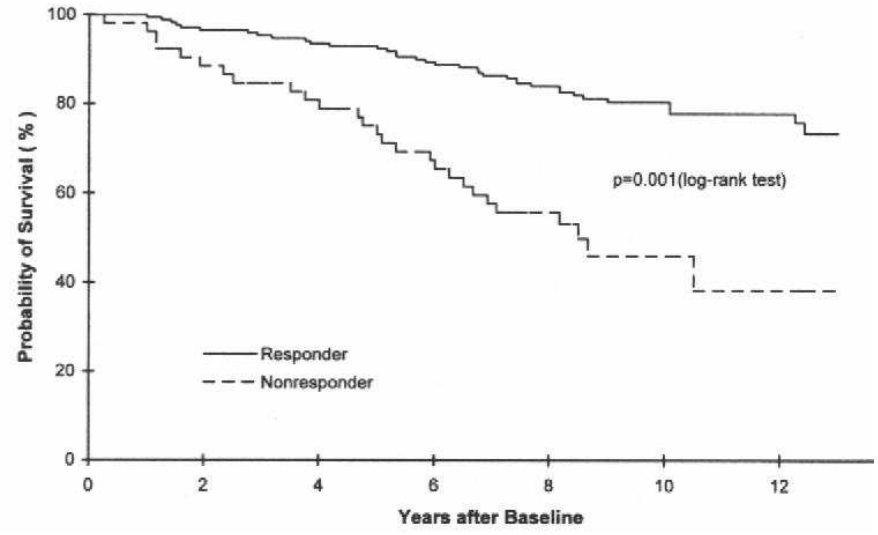
Par ailleurs, plusieurs études ont montré une réduction de l'espérance de vie, estimée à 10 ans, chez les personnes atteintes de PR (Figure 7) (10–12).

**Figure 7 :** Impact de la PR sur la survie, adapté de (10)




En fait, cette réduction concerne surtout les patients les plus graves, réfractaires aux traitements de la maladie (Figure 8) (13).

**Figure 8 :** Surmortalité des patients ne répondant pas au traitement de la PR, d'après (13)




Cette surmortalité associée aux formes graves de PR est due à une augmentation de la prévalence des maladies cardiovasculaires, l'inflammation chronique étant un facteur de risque d'athérome et surtout de déstabilisation des plaques vasculaires athéromateuses (14–17).

Une méta-analyse de 24 études observationnelles utilisant des taux de décès standardisés, a mis en évidence une augmentation de 48% du risque de mortalité cardio-vasculaire chez les patients souffrant de PR par rapport à la population générale (18).

Les mêmes auteurs ont montré dans une méta-analyse très récente de 14 études observationnelles comportant un groupe comparateur, une augmentation de 48% du risque de mortalité cardio-vasculaire incidente chez les patients souffrant de PR par rapport à la population générale (19).

De plus, cet excès de mortalité cardio-vasculaire a aussi été mis en évidence chez des femmes souffrant de polyarthrites récentes, facteur rhumatoïde positif, avec un taux de mortalité standardisé de 2,02 [IC 95% : 1,15-3,2], indépendamment des facteurs de risque cardiovasculaire traditionnels (15). Une étude a montré qu'il existe très tôt dans l'évolution d'une PR, une activation endothéliale et un épaississement de l'intima (augmentation du rapport intima-média) (20).

Les traitements de la PR, en luttant contre l'inflammation chronique, permettent de diminuer cette surmortalité cardiovasculaire (21,22).

## I.4 – Les stratégies thérapeutiques dans la PR

### I.4.1 – *Les différents traitements de la PR*

On dispose maintenant de plusieurs traitements de la PR (Tableau 2), parmi lesquels on distingue :

- Les traitements symptomatiques, qui permettent d'agir au coup par coup, à court terme, sur les symptômes gênants de la maladie ;
- Les traitements de fond, qui ont une action à moyen et long terme et permettent de contrôler l'activité de la maladie (capacité à prévenir les poussées). Certains d'entre eux ont également démontré leur capacité à ralentir, voire arrêter, les destructions articulaires.

A côté des traitements de fond dits synthétiques (23), ces dix dernières années ont vu émerger plusieurs biothérapies qui ont révolutionné le traitement de la PR : les anti-TNF (infliximab, etanercept, adalimumab, certolizumab et golimumab), un inhibiteur de l'interleukine 1 (anakinra), un anti-lymphocyte B (rituximab) et un anti-CD28 (abatacept) (Tableau 2). Mais ces traitements biologiques coûtent chers.

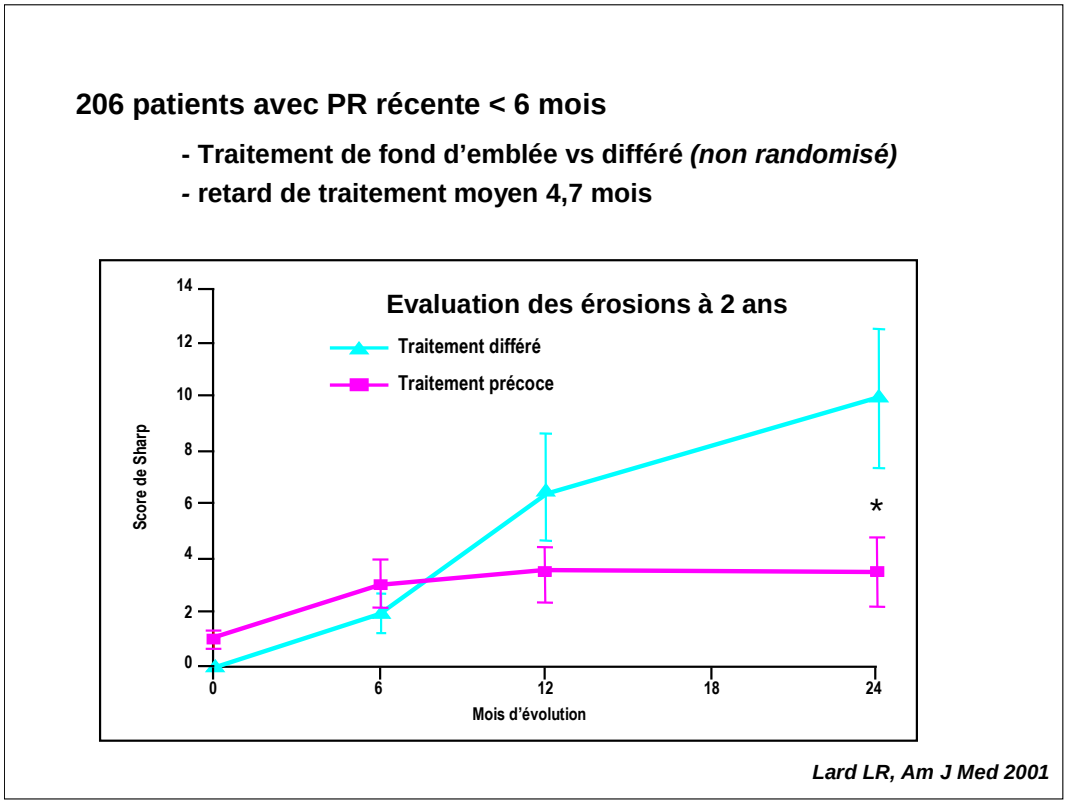
**Tableau 2 :** Les différents traitements de la polyarthrite rhumatoïde.

| **Traitements symptomatiques** | **Traitements de fond** | | |
|---|---|---|---|
| ***Diminution de la douleur et des gonflements à court terme*** | ***Contrôle de l'activité de la maladie (prévention des poussées)*** | ***Contrôle de l'activité de la maladie + Action structurale (prévention des destructions)*** | |
| | Tt de fond synthétique | Tt de fond synthétique | Tt de fond biologique |
| Antalgiques<br>AINS<br>Corticoïdes<br>Infiltrations | Hydroxychloroquine<br>Sels d'or | Méthotrexate<br>Leflunomide<br>Sulfasalazine | Anti-TNFα<br>Anti-CD20<br>CTLA4-Ig<br>Anti-IL6R |

I.4.2 – *La notion de fenêtre thérapeutique*

Le traitement précoce est déterminant dans le succès de la prise en charge de la PR: l'instauration rapide d'un traitement efficace peut permettre d'obtenir une rémission. Cette notion de fenêtre thérapeutique (24) est largement étayée dans la littérature que ce soit au travers d'études observationnelles (25,26), d'essais contrôlés randomisés (27) ou de revues systématiques (28). Un traitement de fond instauré précocement dans la polyarthrite débutante donne de meilleurs taux de réponse, moins d'incapacité fonctionnelle et moins d'atteinte structurale (27,29,30). Ainsi, un traitement de fond débuté dans les 3 premiers mois de la maladie permet d'obtenir une rémission clinique de la maladie chez près de 50% des patients alors que l'initiation du même traitement plus de 6 mois après le début de la maladie ne permet la rémission que dans moins de 10% des cas (31,32). L'effet sur la prévention des destructions articulaires est également plus important lorsque le traitement est débuté précocement (Figure 9) (33).

**Figure 9 :** La notion de fenêtre thérapeutique- traiter tôt, meilleure efficacité structurale (33)




## I.5 – Impact économique de la PR

### I.5.1- *Contexte économique de la PR*

De nombreuses études sur les coûts de la PR ont été réalisées ces dernières années : elles donnent des résultats assez différents. Cela est dû aux caractéristiques de la population étudiée (taille, caractéristiques de la PR, durée d'évolution de la maladie…), à l'organisation du système de soins et surtout à la méthode de calcul des coûts.

Une revue systématique de la littérature a été réalisée en 2009 dans le but de dégager une estimation plus homogène des coûts de la PR (34). Ont été inclues toutes les études (essais contrôlés randomisés et études observationnelles) s'intéressant aux coûts de la maladie d'un point de vue sociétal en Europe occidentale. En tout, 36 études ont été analysées, la plupart dans la PR établie. Le coût direct moyen de la PR (exprimé en euros 2006) était estimé à 4170 € [interquartile range IQR 2756–4561]. Le coût indirect moyen était de 1441 € [IQR 702–1307] selon la méthode dite des «coûts de friction » et de 8452 € [IQR 4144–11566] selon la méthode du capital humain.

La variation dans les estimations des coûts indirects vient principalement de la méthode de calcul. La méthode « des coûts de friction » prend en compte les périodes d'arrêt de travail à court terme qui sont importantes dans la PR débutante. Cependant elle ignore les coûts liés à un arrêt maladie prolongé et à l'incapacité fonctionnelle qui sont prépondérants dans la PR établie (35–37). Ainsi la méthode des « coûts de friction » sous-estime vraisemblablement l'impact réel de la PR sur la perte de productivité et les coûts indirects. Cette revue systématique rapportait également les coûts directs non liés aux soins médicaux qui étaient estimés à 2284 € [IQR 628–3092]. Ils sont principalement le fait des services à domicile et des aides apportées par la famille et les amis.

Cette analyse met en exergue 2 messages :

1) D'un point de vue général, les coûts directs liés à la PR s'élevaient autour de 4000 €, ce qui est cohérent avec les autres revues non systématiques de littérature déjà effectuées sur ce sujet (38,39).

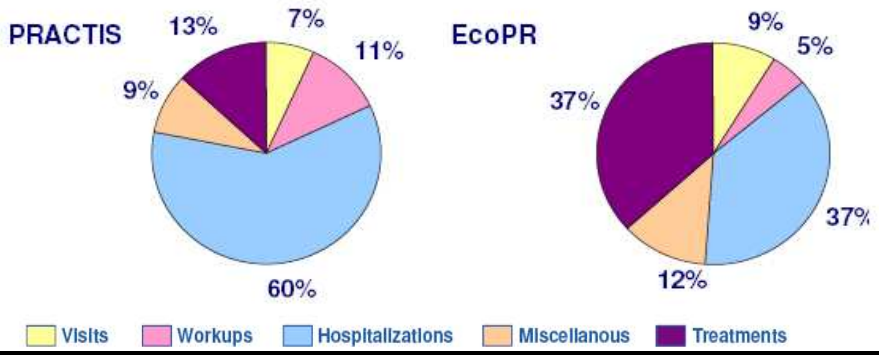
2) Il n'existe pas de consensus sur la méthode de calcul des coûts indirects, ce qui empêche toute comparaison valide (40–43). Ce point est critique pour l'évaluation des maladies chroniques, dans lesquels le retentissement socio-professionnel est important.

### I.5.2- *Evolution des coûts de la PR ces dernières années*

Durant la dernière décennie, les coûts de la PR ont évolué notamment en rapport avec l'émergence des biothérapies. Bien qu'une telle tendance soit difficile à estimer avec précision du fait des difficultés méthodologiques vues plus haut, 2 études observationnelles françaises conduites à 5 ans d'intervalle avec une méthodologie assez proche permettent de tirer certaines conclusions. La 1° étude, l'étude PRACTIS a été conduite dans plusieurs services de rhumatologie en 2000 et a permis d'établir le coût de la PR en France juste avant le lancement des biothérapies sur le marché (44): 1109 patients avec une PR établie ont été inclus et leur consommation de soins a été déterminée sur 1 an. La 2° étude, Eco-PR, a été réalisée parmi les membres d'associations de patients en 2005 sur la consommation de soins chez 1487 patients avec une PR établie ; 27 % étaient traités par anti-TNF (45). La comparaison des 2 études a mis en évidence une très nette augmentation des coûts de la PR sur 5 ans : 4 003 € dans practis versus 11 757 € dans Eco-PR (€ actualisé 2005) (Figure 10).

**Figure 10 :** Evolution des coûts de la PR dans 2 études observationnelles de méthodologie similaire (46)

| | **PRACTIS** (44) | **Eco-PR** (45) |
|---|---|---|
| **N** | 1 003 PR établies | 1 487 PR établies |
| **Biothérapies anti-TNF** | 0 (0%) | 401 (27%) |
| **Coûts directs** | Moy. 4 003 € (med. 2 021 €) | Moy. 11 757 € (med. 3 951 €) |

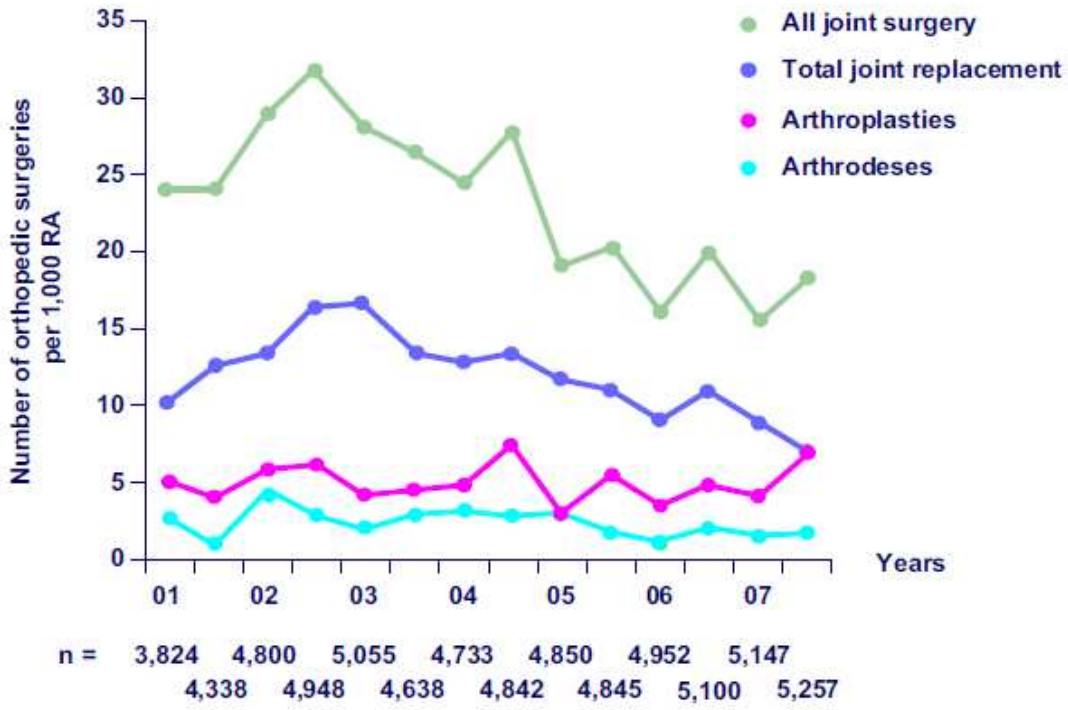



De plus, la répartition des différents coûts a été profondément modifiée avec une diminution des coûts liés à l'hospitalisation mais dans le même temps une forte augmentation des coûts liés aux médicaments due à l'utilisation des biothérapies. Malgré cela, ces médicaments ont été largement acceptés en Europe de part leur efficacité importante qui laisse espérer une diminution des coûts à plus long terme en diminuant le recours aux soins et la perte de productivité.

*Réduction du recours à la chirurgie prothétique*

Une étude dans le registre japonais IORRA a mis en évidence une diminution des interventions chirurgicales orthopédiques dans la PR depuis 2003, date où le premier traitement biologique a été disponible dans ce pays (Figure 11) (47).

**Figure 11 :** Incidence des différents types de chirurgies dans la PR entre 2001 et 2007 dans la cohorte japonaise IORRA, adapté de (47)



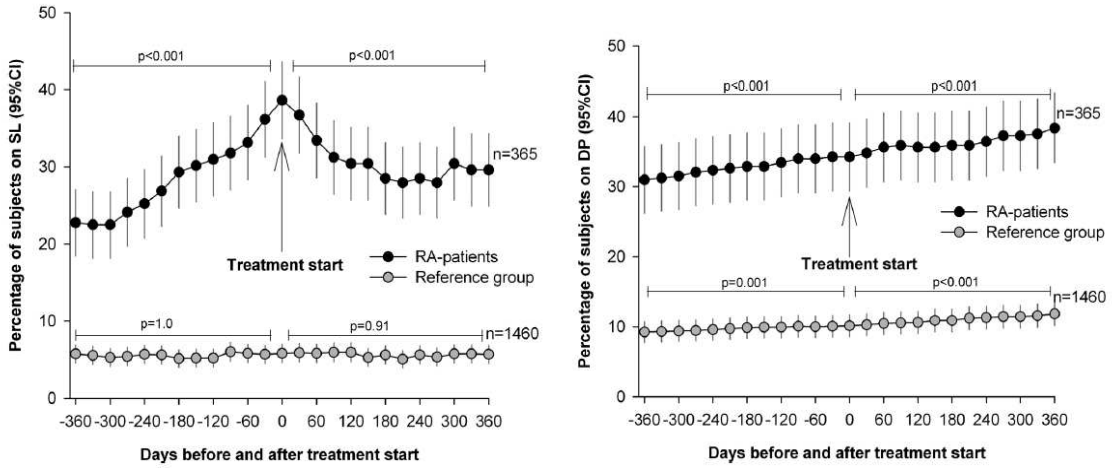

*Réduction de l'impact professionnel*

En raison de la forte relation entre activité de la maladie et incapacité fonctionnelle d'une part et productivité au travail d'autre part, une amélioration substantielle de l'impact de la PR sur le travail était attendue avec les biothérapies. Or une étude internationale QUEST-RA a montré que la probabilité de continuer à travailler chez les patients ayant débuté leur maladie dans les années 2000 était de 80% à 2 ans et de 68% à 5ans (48). Il y avait de fortes disparités selon les pays avec des patients toujours en activité dans certains pays alors que dans d'autres, des patients avec une activité similaire ou même inférieure étaient déclarés en invalidité. Ces différences peuvent s'expliquer par des différences dans le système de protection sociale, le type de profession (manuel…) et l'accès aux soins incluant le recours aux biothérapies.

Une étude très récente a étudié l'évolution de l'incidence de la mise en invalidité chez les patients souffrant de PR entre 1990 et 2010 en Suède (49). Cette incidence diminuait ces dernières années coïncidant avec l'utilisation de traitements plus précoces et plus agressifs par traitements de fond synthétiques et biologiques. Cette diminution des mises en invalidité était présente dans toutes les maladies mais elle était plus marquée dans la PR : en 1990, la PR

représentait 1,9% de toutes les causes de mise en invalidité, elle diminuait à 1,5% en 2000 et à 1% en 2009. Les explications potentielles étaient nombreuses, en dehors d'un effet des traitements : une évolution de la structure d'âge de la population (vieillissement et donc réduction de l'invalidité au profit de la retraite), l'augmentation du niveau d'éducation ou la diminution de la demande physique dans les différents métiers, un éventuel changement dans les politiques de santé (49). Dans le même pays, un autre travail a précisément ciblé l'évolution de la productivité (arrêts-maladie et invalidité) chez les patients avant et après mise sous biothérapie (50). Dans ce travail, les arrêts-maladie augmentent régulièrement dans l'année précédant la mise sous biothérapie, pour ensuite décroître après l'introduction de celle-ci, témoin du meilleur contrôle de la maladie (Figure 12). En revanche, on n'observe aucune inflexion dans la courbe de mise en invalidité, la progression se poursuivant avec la même cinétique dans l'année qui suit l'introduction de la biothérapie, témoignant du fait que la mise en invalidité du fait de la PR n'est pas un choix que médical, mais également social.

**Figure 12 :** Evolution des arrêts-maladie (SL pour *sick leaves*) et des mises en invalidité (DP pour *disability pension*) dans l'année qui précède et qui suit la mise en route d'une biothérapie (50)

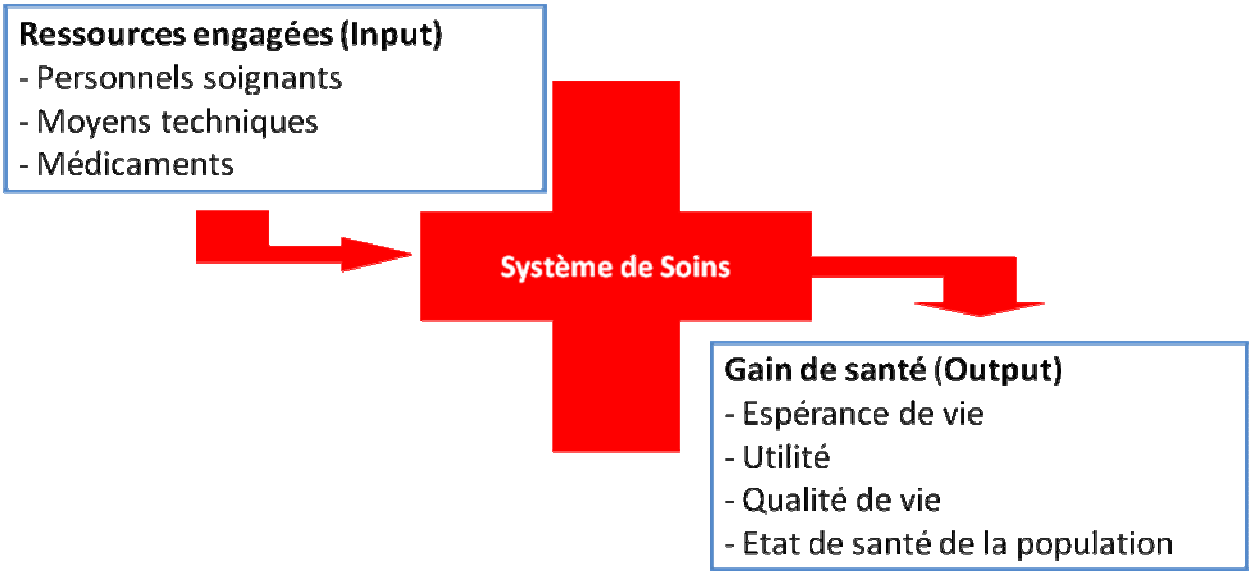


## II – L'EVALUATION MEDICO-ECONOMIQUE

### II.1 – Principe

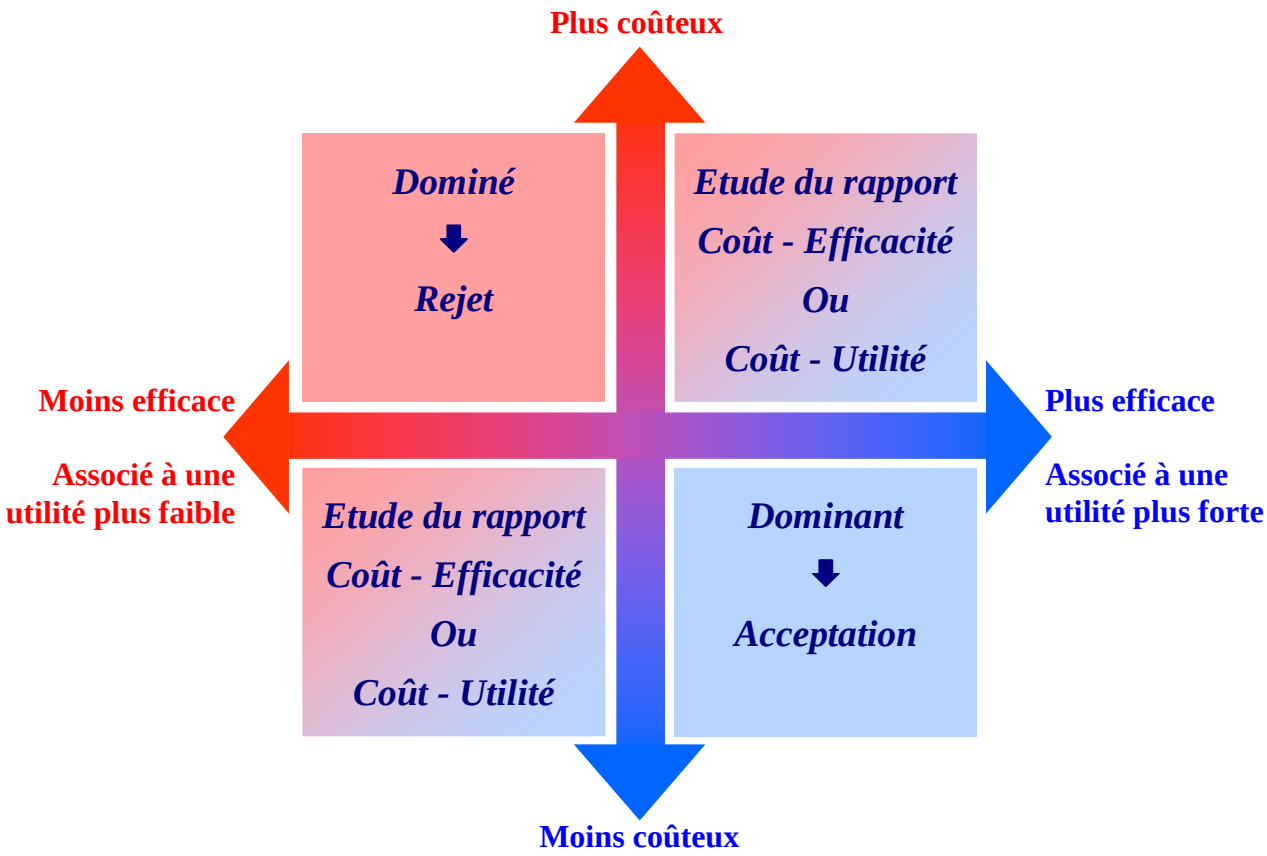
Actuellement, la plupart des systèmes de santé sont confrontés à un problème de financement, lié à une augmentation des dépenses qui deviennent supérieures à la croissance générale de l'économie. Cette problématique implique la mise en balance des ressources mobilisées (*inputs*) et des résultats obtenus (*outputs*), c'est-à-dire des coûts et des bénéfices en termes de santé, afin, à terme, d'établir des priorités et de formuler des choix (Figure 13) (51). L'objectif des études médico-économiques comparatives, ou études coûts – avantages (52,53) est d'optimiser la santé de la population à dépense égale ou d'optimiser les dépenses à niveau de santé constant.

**Figure 13 :** Mise en opposition des inputs et des outputs.



A partir de modélisation, utilisant souvent le modèle à chaîne de Markov, il est possible de mettre en balance les *inputs* et les *outputs* pour deux options (54). On peut ainsi comparer dans un même temps l'efficacité et le coût (Figure 14).

**Figure 14 :** Aide à la décision des études médico-économiques comparatives




Dans 2 types de situations, le choix est aisé. Lorsqu'une option A engage plus de ressources et apporte moins de gains pour la santé – en d'autres termes, est plus coûteuse et est moins efficace – qu'une option B, on conclut au caractère dominé de l'option A : le développement de l'option A n'a pas d'intérêt. A l'opposé, lorsque cette option A apporte plus de gains tout en mobilisant moins de ressources – c'est-à-dire est plus efficace tout en étant moins coûteuse –, on conclut à son caractère dominant, et l'option A doit être développée et mise en pratique.

Ces situations sont cependant rares et, dans la majorité des cas, un gain en efficacité est contre balancé par un coût supérieur. La difficulté dans ces situations est de déterminer ce qui peut être considéré comme acceptable, c'est-à-dire jusqu'à quel point on peut accroître les ressources dédiées à la santé pour améliorer l'état de santé de la population.

### II.2 – Typologie des coûts

Les coûts d'une maladie (*Cost of Illness*) permettent de juger de l'impact économique d'une maladie (52,55). Ils peuvent être démembrés en 3 groupes (41,52).

- Les coûts directs, correspondant à la prise en charge médicale et paramédicale de la maladie ;
- Les coûts indirects, traduisant la perte de productivité d'un individu liée à une maladie, à ses conséquences ou à sa prise en charge ;
- Les coûts intangibles, tentant de traduire en termes monétaires l'impact humain d'une maladie (douleur, handicap, retentissement psychologique).

#### II.2.1 – *Coûts directs*

Les coûts directs regroupent tous les éléments de la prise en charge médicale et paramédicale des patients : consultations, hospitalisations, traitements avec leur surveillance et la gestion d'éventuels effets indésirables, soins paramédicaux, etc... Ces frais sont, en général, pris en charge par les différents régimes d'assurance-maladie, publics ou privés, mais certains restent à la charge des patients.

Leur mesure pour des pathologies aiguës peut reposer sur des questionnaires remplis par les patients ou sur des bases de données d'assurances ou d'hôpitaux. Pour les pathologies chroniques, les auto-questionnaires sont les plus souvent utilisés pour recueillir les informations relatives à la consommation de soins et aux dépenses effectuées pour se soigner. Ils couvrent en général une période de 1 à 6 mois. Dans certains cas, les informations pour la détermination de ces coûts sont issues d'essais thérapeutiques contrôlés, et parfois de bases de données de l'assurance maladie.

Ces estimations ont cependant des limites: problème de mémorisation (oubli des patients), imprécision des réponses données (par exemple, « scanner » sans mention du site exploré, ou « prise de sang » sans spécification des tests prescrits), variations des prix (prix de journée d'hospitalisation différents selon l'établissement et sa localisation, prix de consultation flexibles selon le secteur d'activité, prix des médicaments fluctuants selon le lieu de délivrance, en hôpital ou en officine de ville).

### II.2.2 – *Coûts indirects*

Les coûts indirects sont liés à la perte de « productivité » engendrée par une maladie, qu'elle soit aiguë ou chronique. Le terme de « productivité » ne se limite pas ici à la seule capacité d'un individu à exercer une activité professionnelle. Il englobe toutes les activités ou fonctions qu'une personne peut remplir dans sa vie courante et qui lui permettent de tenir son rôle dans la société : travail rémunéré, tâches ménagères, rôle d'éducation, aide à une tierce personne, bénévolat, etc. La notion de « perte de productivité » peut donc concerner tous ces domaines : changement de poste du fait de la maladie, diminution des perspectives de carrière, réduction d'activités de bénévolat, difficultés pour l'entretien de sa maison ou l'éducation des enfants, etc.

La détermination des coûts indirects est donc plus complexe que celle des coûts directs. En effet, il est difficile de savoir ce qu'une personne aurait gagné en l'absence d'une maladie (cas principalement des maladies chroniques), ou de valoriser des tâches habituellement non rémunérées telles qu'élever un enfant pour ne citer qu'un exemple.

Il existe différentes techniques de valorisation (transformation en unité monétaire) de la perte de productivité qui peut être limitée aux seules activités rémunérées, sur une durée maximale de 3 mois dans le cadre dite des « coûts de friction » (au-delà de cette période, le travailleur malade est mis en longue maladie et remplacé par une autre personne puisqu'on n'est pas dans une situation de plein emploi) ou bien étendue aux activités non rémunérées (bénévolat par exemple), sans limite de temps dans la méthode dite du « capital humain ». Cependant aucune ne fait consensus.

### II.2.3 – *Coûts intangibles*

Les coûts intangibles correspondent aux conséquences humaines d'une maladie, telles que le handicap, la douleur physique ou le retentissement psychologique d'une maladie. En fait, ils visent à quantifier en termes monétaires l'altération de la qualité de vie.

Cependant, étant donné les difficultés de valoriser de telles notions, la majorité des études évaluant les coûts d'une maladie n'aborde pas les coûts intangibles.

## II.3 – Typologie des gains de santé

Dans la perspective de définition de priorités, comme il est nécessaire d'estimer les coûts, il est indispensable de mesurer les gains de santé. Il existe plusieurs façons d'y parvenir expliquant les différents types d'études coûts – avantages (Tableau 4) (52,53,56–58).

**Tableau 4 :** Les différents types d'études coûts – avantages.

| | **Evaluation des effets (efficacité ou tolérance)** | |
|---|---|---|
| | **Gains de santé valorisés** | **Unité de mesure** |
| **Coût – bénéfice** | Evolution de l'état de santé | Unité monétaire ( €, $, …) |
| **Coût – efficacité** | 1 effet clinique unique (ex : douleur, survie…) | Unité « naturelles » (ex : diminution de la douleur sur l'EVA, années de vie gagnées,…) |
| **Coût – utilité** | Evolution de l'état de santé | QALY |

II.3.1 – *Une vision comptable : gain exprimé en unité monétaire*

Il est potentiellement intéressant d'exprimer un gain ou un avantage (output) dans le domaine de la santé directement en unité monétaire. Ceci est a priori difficile puisqu'il s'agit alors de mettre une valeur sur une valeur immatérielle – la santé -, par nature non monnayable : par exemple, le prix d'une vie ou le prix d'une parfaite santé. Il existe pour arriver à cela des techniques, notamment celles dites de valorisation contingente (propension à payer etc…).

En théorie, ces études coûts – avantages présentent l'avantage de pouvoir comparer tout type de programmes ou d'options quelque soit le domaine (santé, environnement etc…), puisque l'unité de mesure est universelle. Cependant ce type d'études est très peu utilisé en santé.

II.3.2 – *La vision des médecins: gain d'efficacité exprimé en unité naturelle*

La quantification des gains ou des bénéfices que l'on peut attendre d'un traitement ou d'une prise en charge peut se faire de différentes façons :

- soit en unités dites « naturelles », c'est-à-dire spécifiques d'une maladie : la douleur sur l'EVA, l'activité de la maladie avec le nombre d'articulations gonflées/douloureuses, la VS ou le DAS28 ;
- soit en unités plus généralistes : l'incapacité fonctionnelle avec l'exemple de l'indice HAQ, la qualité de vie mesurée par des outils génériques (SF-36) ou plus spécifiques de groupe de maladies (Emir-court/AIMS2-SF)

Un gain en santé peut être ainsi exprimé en termes d'efficacité: par exemple, fréquence des rémissions ou le nombre d'années de vie sans rechute pour un traitement anti-néoplasique, douleur ou incapacité fonctionnelle pour une thérapeutique rhumatologique, fractures évitées pour un traitement anti-ostéoporotique etc...

Comme on peut le comprendre, ces études permettent de comparer des traitements ayant une action similaire ou proche. Elles ne permettent pas de déterminer une préférence entre deux programmes radicalement différents, par exemple un programme de prévention des lombalgies chroniques et un traitement de l'accident vasculaire cérébral ; l'un des programmes vise à diminuer les douleurs et à maintenir ou améliorer les capacités fonctionnelles des patients concernés, l'autre vise à traiter et éviter les accidents vasculaires et le risque de mort qui en découle.

### II.3.3 – *Gain d'utilité: un nécessaire compromis*

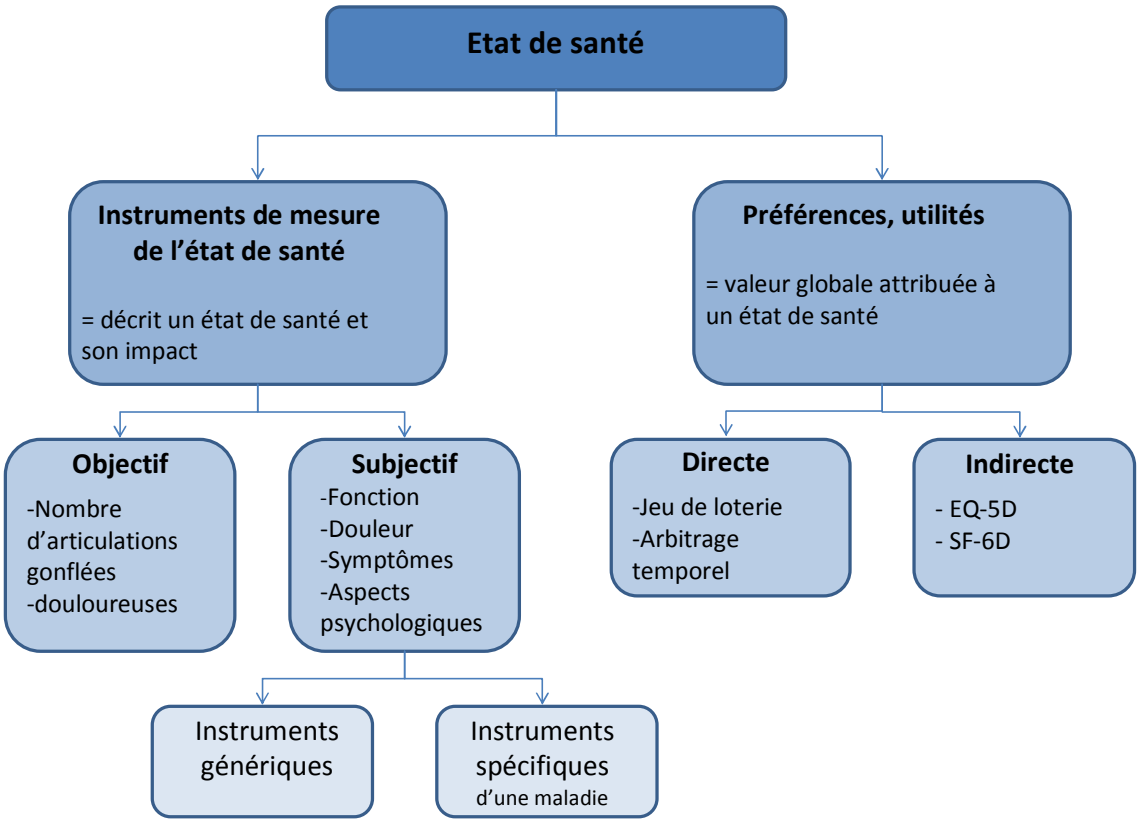
La notion d'utilité est une mesure permettant d'attribuer une valeur numérique à un état de santé, variant de 0, a priori la mort, à 1, un état de parfaite santé.

En réalisant des mesures répétées au cours du temps, on peut estimer l'évolution de l'utilité et combiner la quantité et la qualité de vie. On peut alors exprimer cela en Année de Vie Ajustée à la Qualité, ou QALY (*Quality Adjusted Life Year*).

## III – LA NOTION D'UTILITE

Depuis quelques années, en complément des traditionnels indices de morbidité et de mortalité, de nouveaux instruments de mesure de l'état de santé ont été développés. Si l'évaluation des interventions en santé est restée très longtemps basée sur la mise en balance du bénéfice purement clinique pour les malades au regard des effets indésirables potentiels, la mesure de la qualité de vie liée à la santé est devenue une dimension essentielle (Figure 15). La quantification du gain de santé est un enjeu important, notamment dans le contexte actuel avec des thérapeutiques de plus en plus coûteuses et des ressources limitées.

**Figure 15 :** Description des différents instruments de mesure des gains de santé, adapté de (59)



## III.1 – Principes

### III.1.1 – *Les préférences individuelles en santé*

Le concept d'utilité vient de la philosophie utilitariste et a notamment été développé dans la recherche sur les processus de décision. La théorie de l'utilité espérée a été développée par Von Neumann et Morgenstern dans leur ouvrage de 1944 sur la théorie des jeux (60). Les auteurs appliquent cette théorie pour prédire le comportement des joueurs dans les jeux non coopératifs. Cette théorie a pour but d'expliciter les choix des individus et leurs préférences pour différentes situations en univers incertain. Elle repose sur le principe selon lequel l'utilité d'une perspective aléatoire est égale à la somme des utilités des termes de l'alternative pondérées par leur probabilité respective.

En économie, les individus sont supposés avoir le choix entre des actions dont les conséquences sont évaluées par leur utilité. Celle-ci est assimilable au niveau de satisfaction que procure le fait d'en disposer. Lorsque les conséquences sont incertaines, l'individu "rationnel" choisit l'action dont l'utilité espérée est la plus grande. Torrance définit l'utilité comme "le désir ou la préférence d'un individu pour une situation" (61).

La théorie de l'utilité espérée de von Neumann et Morgenstern repose sur 3 axiomes fondamentaux :

- la transitivité → reformulée de la façon suivante par Bell et Farquhar (62) : pour toute paire de loteries y et y', ou bien y est préféré à y', ou bien y' est préféré à y, ou bien l'individu est indifférent entre y et y'. De plus pour tout triplet de loteries y, y' et y' ', si y est préféré à y' et si y' est préféré à y' ', alors y est préféré à y' '. De même si y est indifférent à y' et si y' est indifférent à y' ', alors y est indifférent à y' '
- l'indépendance → la valeur donnée à un résultat est indépendante de la manière dont il est obtenu ou de son contexte
- et la continuité des préférences → si on a trois résultats x1, x2 et x3, tels que x1 est préféré à x2 qui est préféré à x3, il existe une probabilité p pour laquelle l'individu est indifférent entre le résultat x2 obtenu avec certitude et la loterie qui correspond au résultat x1 avec la probabilité p et au résultat x3 avec la probabilité 1 – p (62).

À l'origine, l'utilité est un concept purement empirique, qui s'apparentait à un indicateur de santé à la fois unidimensionnel et générique. Il synthétisait sous un chiffre unique les dimensions multiples de la qualité de vie (douleur, mobilité, capacités cognitives, relations sociales, etc.) et il était susceptible de s'appliquer à tous les contextes médicaux.

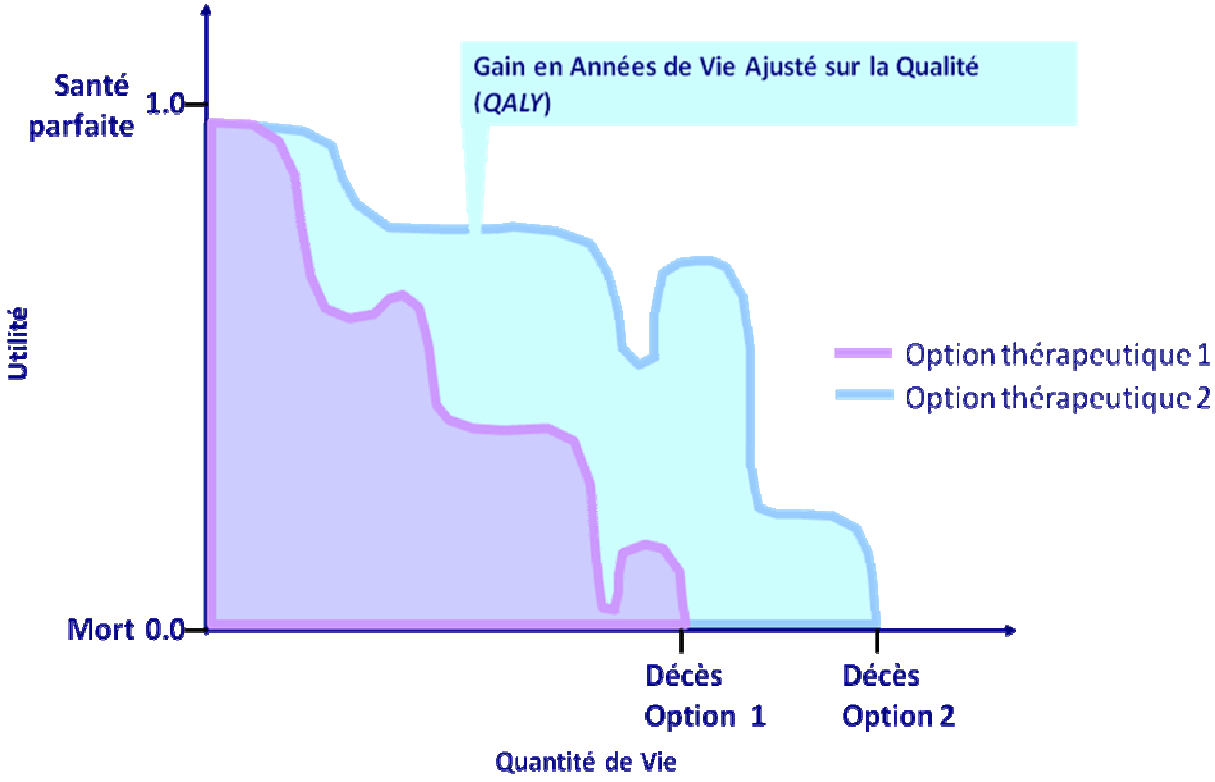
A l'inverse, la plupart des échelles médicales de mesure d'états de santé sont multidimensionnelles et parfois spécifiques d'une pathologie.

Le concept d'« utilité » est donc proche de celui de qualité de vie et a principalement été développé par des économistes de la santé. La notion d'utilité est une mesure permettant d'attribuer une valeur numérique à un état de santé, variant de 0, à priori la mort, à 1, un état de parfaite santé. Il repose sur l'expression, par un répondant ou un groupe de répondants, d'une préférence révélée entre deux ou plusieurs états de santé (*revealed preference*). Lorsque le répondant est indifférent entre 2 états de santé (pas de préférence ou préférence nulle), les 2 états de santé sont associés à une utilité identique.

### III.1.2 – *Le QALY*

En réalisant des mesures répétées de l'utilité au cours du temps, on peut combiner la quantité et la qualité de vie. L'aire sous la courbe représente une utilité combinée au temps que l'on exprime en Année de Vie Ajustée à la Qualité, ou QALY (Quality Adjusted Life Year) (Figure 16). Ce concept a été introduit pour mesurer à la fois la quantité (durée de vie restante ou espérance de vie) et la qualité de vie (63–65). Il consiste à pondérer le temps passé dans un état de santé donné par un coefficient (compris entre -1 et 1) rendant compte de la qualité de vie liée à un état de santé et du niveau d'utilité de l'état de santé considéré. Par définition, un poids de 0 est affecté à la mort alors qu'un poids de 1 reflète la pleine santé. Une année dans un état de parfaite santé équivaut à deux années dans un état de santé dont l'utilité est diminuée de moitié.

**Figure 16 :** Principe des QALY




Grâce aux QALY, il est donc possible (en théorie) de comparer toutes les actions de santé quelque soit le traitement, quelque soit la maladie.

### III.2 – Méthodes d'obtention de l'utilité

Il existe différentes façons de procéder pour déterminer l'utilité associée à un état de santé.
On distingue deux types d'approches (65) :

- L'approche directe qui consiste à demander aux patients inclus dans les études de valoriser directement leur propre état de santé à l'aide de différentes méthodes : principalement jeu de loterie (*standard gamble*) ou arbitrage temporel (*time trade-off*).
- L'approche indirecte qui consiste à utiliser des « systèmes de classification des états de santé multi-attributs » ou « index d'utilité ». Ils permettent de définir l'état de santé de la personne interrogée à l'aide d'un questionnaire standardisé et de lui associer une valeur d'utilité qui est celle affectée par la collectivité à cet état.

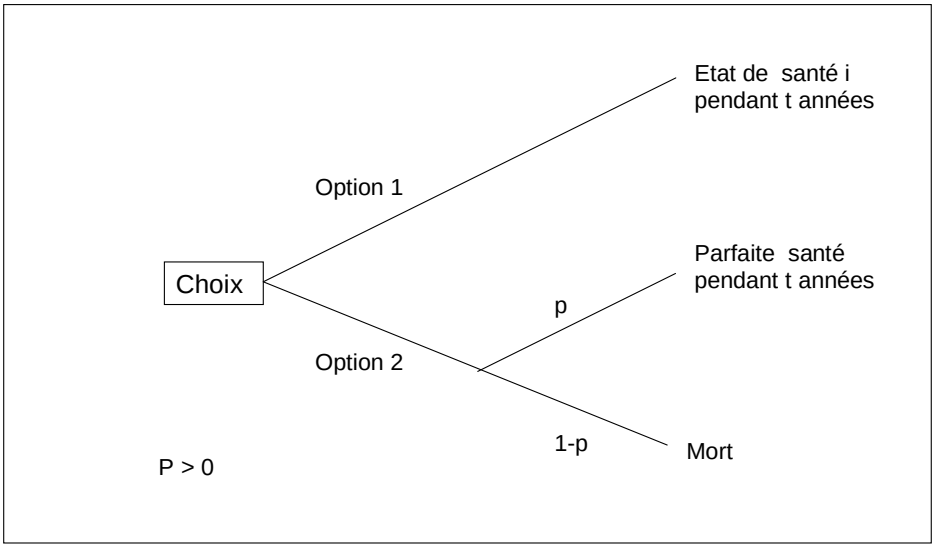
### III.2.1 – *Utilité directe*

Plusieurs méthodes, appelées « méthodes de révélation des préférences », permettent de calculer les niveaux d'utilité. Il s'agit du jeu de loterie ou de l'arbitrage temporel qui sont les deux outils les plus couramment employés et qui répondent aux exigences méthodologiques sous-tendant la réalisation des analyses de type coût-utilité. En effet, l'échelle numérique (rating scale) n'est pas au sens strict une mesure d'utilité car elle n'implique pas d'incertitude, de risque ou de sacrifice. Ces deux techniques sont réalisées en entretien individuel « face à face » avec le patient, assisté par un ordinateur.

#### III.2.1.1 – Jeu de loterie

Avec le jeu de loterie, deux contrats sont proposés au patient : le premier est d'accepter de vivre avec certitude dans un état de santé donné ; le second comporte deux résultats possibles: une parfaite santé pendant le reste de sa vie, avec une probabilité p de décéder immédiatement. On fait varier la probabilité p jusqu'à ce que la personne soit indifférente aux deux options. Cette valeur est celle de la préférence attribuée à l'état de santé étudié (Figure 17). Le score d'utilité obtenu par le jeu de loterie varie entre 0 (la mort) et 1 (un état de parfaite santé). Par exemple, une utilité de 0,7 indique que le patient est prêt à accepter jusqu'à 30% de risque de mort immédiate en échange d'une santé parfaite ([1-0,7]*100). Le jeu de loterie est classiquement considéré comme la méthode de référence pour les mesures de préférences car l'on raisonne en contexte d'incertitude avec la notion de risque (66,67).

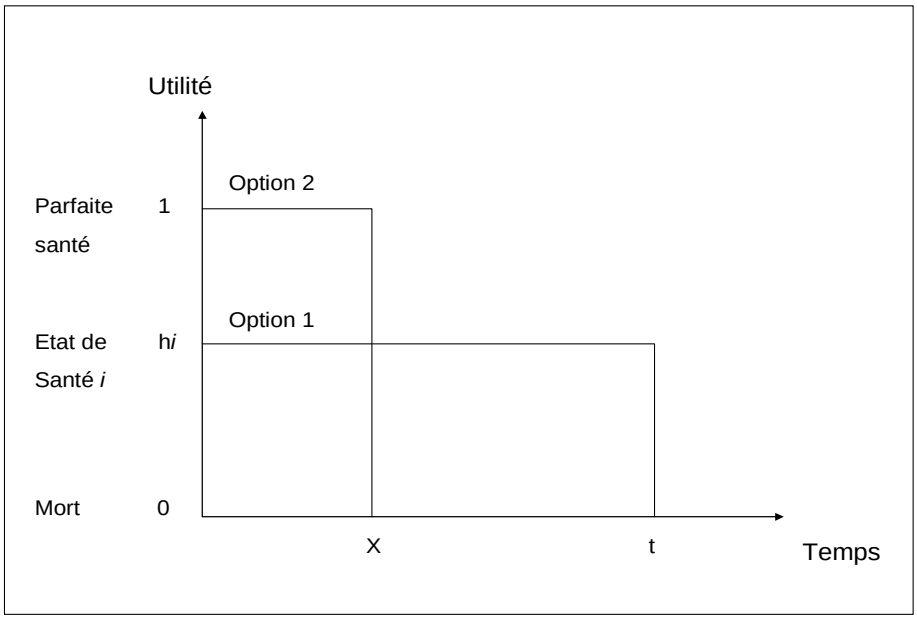
**Figure 17 :** Principe du jeu de loterie d'après (68)



### III.2.2.2 – Arbitrage temporel

Avec l'arbitrage temporel, deux options sont proposées au patient : la première est de vivre dans son état de santé pendant une durée égale à l'espérance de vie d'un individu dans cet état (t), puis de mourir ; la deuxième est de vivre en parfaite santé pendant une durée plus courte (x), puis de mourir. On fait varier le temps x jusqu'à ce que la personne interrogée soit indifférente aux deux options, point auquel le score cherché de la préférence pour l'état de santé étudié est donné par le rapport x/t (Figure 18). Une utilité de 0,7 en arbitrage temporel indique que le patient est prêt à diminuer son espérance de vie de 30% en échange d'un état de santé parfaite.

**Figure 18 :** Principe de l'arbitrage temporel d'après (68)




Certains auteurs considèrent que l'arbitrage temporel décrit mieux les préférences individuelles car il évite le problème de l'aversion au risque et est plus facile à comprendre pour les répondants (69). La notion de temporalité au risque intervient également: immédiat dans le jeu de loterie, différé dans l'arbitrage temporel (70). Dans les maladies où le risque est à très court terme (cancer…), les patients tolèrent mieux le jeu de loterie que dans les maladies chroniques où le risque est plus tardif.

### III.2.2 – *Utilité indirecte*

De façon plus récente, des auto-questionnaires ont été développés, à la fois plus simples d'utilisation et plus acceptables pour les répondants : EQ-5D, SF-6D ou d'autres (71–73). Les valeurs d'utilité peuvent être alors déterminées indirectement par utilisation de ces questionnaires multidimensionnels. Une valeur d'utilité est attribuée à l'aide d'algorithmes dérivés de valeurs issues de populations de référence. Les utilités directe et indirecte présentent chacune des avantages et des inconvénients (Tableau 5).

**Tableau 5 :** Les différentes mesures d'utilité (59)

| | **Avantages** | **Inconvénients** |
|---|---|---|
| **Utilité directe** | - Fournit un chiffre unique | - Pourrait ne pas être aussi sensible au changement que les instruments de mesure de l'état de santé<br>- Nécessite un entretien individuel en « face à face »<br>- Concept difficile à appréhender pour certains sujets<br>- Influencé par l'attitude du patient vis-à-vis du risque, du temps |
| **Utilité indirecte** | - Fournit un chiffre unique<br>- Plus appropriée pour les autorités de santé car implique des pondérations issues de la population générale<br>- Pas besoin d'entretien individuel<br>- Plus simple (auto questionnaire) | - Pourrait ne pas être aussi sensible au changement que les instruments de mesure de l'état de santé<br>- Risque moins perceptible par la personne qui remplit le questionnaire (biais de complaisance)<br>- Pas de risque encouru |

L'EQ-5D (74) et le SF-6D (72) sont 2 questionnaires multi-attributs de plus en plus utilisés pour estimer l'utilité dans les évaluations médico-économiques. Ils mesurent la perception des individus, mais aussi celle de la société à travers les poids sociétaux (pondérations issues d'enquêtes en population) qui sont affectés à divers états décrits par les individus.

#### III.2.2.1 – EQ-5D

L'EQ-5D est un système descriptif comportant une question pour 5 dimensions : mobilité, autonomie de la personne, activités courantes, douleurs/gêne, anxiété/dépression. Pour chaque item, il existe 3 niveaux de réponses : « aucun problème » (1), « des problèmes » (2), « des problèmes extrêmes » (3). Le répondant doit indiquer, pour chacune des dimensions, l'état dans lequel il se trouve (annexe 1). Les réponses données peuvent être combinées en un nombre de 5 digits décrivant l'état de santé du répondant. Par exemple, une personne ne présentant aucun problème pour se déplacer à pied et pour prendre soin d'elle-même, des problèmes pour accomplir ses activités quotidiennes, qui a des douleurs ou une gêne modérée et qui est extrêmement anxieuse ou déprimée aura un profil d'état de santé noté 11223. Au total, l'EQ-5D permet de décrire 243 ($3^5$) états de santé possibles. Dans de nombreux pays, une fonction d'utilité a été déterminée et permet d'associer un poids à chacun des états de santé définis par l'EQ-5D (74). L'algorithme dérivé des valeurs issues de la population générale française en utilisant la technique de l'arbitrage temporel vient d'être publié (75). L'utilité qui en découle varie entre -0.529 à 1, 1 représentant un état de santé parfaite, 0 la mort et les valeurs négatives, des états de santé "pires que la mort".

Une version de l'EQ-5D comportant 5 niveaux par dimension vient d'être développée par le groupe EuroQol, permettant d'accroître le nombre d'états de santé décrits. Mais les études permettant d'obtenir la valorisation de l'échelle à 5 niveaux ne sont pas encore disponibles. En l'état actuel des choses, la nouvelle version de l'instrument EQ-5D ne peut être insérée dans les questionnaires que dans un but descriptif et ne peut pas être utilisée dans les études économiques.

#### III.2.2.2 – SF-6D

Le SF-36 comporte 36 items et permet d'obtenir un profil de qualité de vie liée à la santé évaluant 8 domaines distincts : Activités physiques, Limitations dues à l'état physique, Douleurs physiques, Vie et relations avec les autres, Santé psychique, Limitations dues à l'état psychique, Vitalité et Santé perçue (76). Il comporte également une question de changement et 2 scores résumés standardisées. On obtient alors des scores par dimension dont le 0 défini la qualité de vie la pire et 100 la meilleure.

Le SF-6D peut se dériver de 11 questions du SF36 en utilisant l'algorithme de Brazier et al ce qui fournit alors un score unique (72). Il comporte 6 dimensions: Fonction physique, Limitation de rôle, Fonction sociale, Douleur, Santé mentale et Vitalité avec 4 à 6 niveaux de réponses selon les dimensions. Le SF-6D permet ainsi de décrire 18 000 états de santé. Une fonction de pondération permettant d'affecter un poids à chacun de ces états a été calculée au Royaume-Uni (72) et dans plusieurs autres pays par la technique du jeu de loterie. Si le questionnaire SF-36 a été adapté et validé en français (77), l'index dérivé n'est, à ce jour, pas disponible avec des pondérations françaises. Le score d'utilité obtenu avec le SF-6D varie entre 0,29 et 1 (72).

**Tableau 6:** Présentation des 2 instruments d'utilité indirecte l'EQ-5D et le SF-6D

| | **Dimensions (niveaux de réponse)** | **N questions** | **N états** | **Méthode** | **Etendue** |
|---|---|---|---|---|---|
| **EQ-5D** | Mobilité (3)<br>Autonomie de la personne (3)<br>Activités courantes (3)<br>Douleur/gêne (3)<br>Anxiété/dépression (3) | 5 questions | 243 | Arbitrage temporel | -0,59 à 1,00 |
| **SF-6D** | Fonction physique (6)<br>Limitation de rôle (4)<br>Fonction sociale(5)<br>Douleur (6)<br>Santé mentale(5)<br>Vitalité (5) | 11 questions du SF-36 | 18 000 | Jeu de loterie | 0,296 à 1,00 |

### III.3 – Avantages de l'utilité et des QALYS

L'utilité présente l'avantage de capturer en une seule donnée chiffrée différents paramètres de nature différente et d'importance inégale. L'approche utilité-préférence permet ainsi de disposer d'un score unique combinant à la fois les bénéfices et les effets indésirables d'une thérapeutique, ainsi que les diverses répercussions notamment physiques et psychologiques de la maladie. Si les conséquences ne peuvent être réduites en un indicateur unique et physique, l'approche coût-utilité, qui tend à devenir la méthode de référence des analyses médico-économiques, consiste à les mesurer par un indicateur, le QALY (quality adjusted life years) qui pondère le temps passé dans un état de santé par un coefficient rendant compte de sa qualité. Certains économistes ont proposé d'intégrer le QALY dans la définition des priorités d'un système de santé, d'autres de l'utiliser pour accepter la diffusion d'une procédure médicale. Toutes ces propositions ont soulevé de nombreuses objections. Les économistes favorables au QALY considèrent cependant qu'il éclaire les décisions d'allocation des ressources et entraîne la réflexion sur la place de l'usager dans le système de soins.

### III.4 – Limites de l'utilité et des QALYS

La notion de QALYs a en effet suscité depuis son apparition de nombreuses critiques souvent venues de sociologues (78) voire des économistes eux-mêmes (79).

Beaucoup trouvent en effet abusive la réduction des états de santé à une dimension unique et l'assimilation de la mort à un état de santé comme les autres.

De plus, plusieurs critiques peuvent être formulées concernant la validité des QALY (79,80) et des hypothèses implicites sous-tendant ce modèle (81).

Voici 5 des hypothèses les plus discutées :

- La qualité de vie est indépendante de la durée de vie. Considérons par exemple, 2 options. Option A : 14 ans de parfaite santé comparés à 50% de chance de vivre 10 ou 20 ans de parfaite santé. Option B : 14 ans avec une maladie chronique versus 50% de chance de vivre 10 ou 20 ans avec cette maladie chronique. Il n'est pas sûr que ce point d'indifférence soit identique dans ces 2 scénarios.

- La relation entre quantité et qualité de vie est constante ; la différence d'utilité entre 40 ans de vie avec un handicap et 32 ans en parfaite santé est égale à celle existant entre 30 ans avec un handicap et 24 ans sans handicap. Il n'est pas certain que ce ratio demeure constant au cours du temps et la perception du court-terme peut différer de celle du moyen et long-terme.

- L'utilité associée à un état de santé est constante au cours du temps. En fait, la perception d'un état de santé par un répondant peut varier selon les circonstances de la vie de ce répondant au moment où il donne la réponse. De même, un effet secondaire survenant dans les 5 ans ne sera pas perçu de la même manière que s'il arrive dans les premiers jours de traitement. Il y a un risque de surestimer l'impact négatif sur la qualité de vie d'un effet secondaire à court-terme.

- La neutralité au risque ; il est considéré que face à un même risque, les individus semblables feront les mêmes choix. Or cette hypothèse ne prend pas en compte le comportement individuel face au risque : certains individus ont une propension au risque (beaucoup d'apnéistes) et d'autres une aversion au risque (beaucoup de professeurs de rhumatologie) (81).

- La moyenne en QALYs d'un groupe est équivalente à la moyenne en QALYs de chaque membre du groupe. Il paraît évident que l'hétérogénéité inter-individu invalide cette hypothèse d'additivité.

D'autres éléments, qui peuvent influencer la mesure de la qualité de vie et des changements d'utilité indirecte, sont à prendre en compte.
Le concept de multimorbidité a été défini par Van Den Akker et al. comme l'existence d'au moins 2 maladies chroniques (82). Dans une revue systématique sur l'impact de la multimorbidité sur la qualité de vie, Fortin et al. ont montré qu'il existait une relation inverse entre ces deux concepts et que la multimorbidité pouvait toucher toutes les dimensions de la qualité de vie (83).

Un autre phénomène à considérer est celui de *response-shift*. Sprangers et Schwartz le définissent comme un changement de l'état de santé d'un individu qui peut modifier sa façon de percevoir le monde, ses propres normes ou sa manière d'appréhender le réel (84). Les

patients atteints d'une pathologie chronique rapportent souvent une amélioration de leur qualité de vie au cours du temps malgré l'absence d'amélioration objective de la maladie, sûrement en rapport avec l'adaptation de l'individu à sa maladie.
De plus, dans l'utilité indirecte, si l'état de santé est bien décrit par le patient par le biais d'auto-questionnaires, les pondérations sont celles de la population générale, donc de personnes indemnes de la pathologie en question.

Le modèle pose par ailleurs des questions éthiques : est-il légitime d'affecter à une vie humaine une valeur inférieure à une autre sous prétexte que sa qualité est dégradée ? La valeur humaine n'est elle pas intangible et sa valeur indépendante d'une « norme » de bonne santé ?

D'autres attirent l'attention sur les injustices possibles en matière d'allocation des ressources, la détermination des QALYs défavorisant les personnes âgées par rapport aux jeunes, les plus malades par rapport aux moins malades, les pathologies chroniques par rapport aux maladies aiguës, les prises en charge préventives par rapport aux soins curatifs, etc. (85). L'effet des mesures de santé sur les générations futures n'est pas non plus pris en compte.

Une controverse a récemment opposé les QALYs à une autre méthode de valorisation des états de santé, les "HealthyYears Equivalent" (HYEs) (86). Mehrez et Gafni reprochent aux QALYs de ne prendre que partiellement en compte les préférences des individus: la théorie de l'utilité n'est en effet utilisée que pour obtenir des coefficients de qualité de vie affectés à un nombre d'années T fixe (87). Les préférences portent sur la qualité de vie q et non sur les couples quantité-qualité (q, t), ce qui permettrait par exemple, de choisir une vie plus courte mais de meilleure qualité que l'état de référence. Ils proposent donc une mesure alternative, les HYEs, qui prend en compte à la fois la durée et la qualité de vie, tout en restant facilement interprétable et utilisable. Les HYEs sont définis comme le nombre d'années de santé parfaite donnant la même utilité que le profil initial QT .
Les HYEs ont l'avantage de donner une mesure intégrée des préférences sur le couple qualité-quantité de vie et de ne pas nécessiter de spécification précise de la fonction d'utilité. Plus général sur un plan théorique, ce concept s'avère en pratique très difficile à utiliser. En effet, les méthodes imaginées par ces promoteurs nécessitent les mêmes hypothèses que celles des QALYs. Ainsi, dispose-t-on à la fois d'un concept plus facile à mesurer mais avec des failles théoriques (QALYs) et d'un concept plus riche théoriquement, mais décevant en pratique

(HYEs) (88). Au-delà des aspects techniques, ce débat a l'intérêt de révéler la portée et les limites du recours à la théorie économique pour représenter les états de santé.

Si la littérature souligne l'insuffisance de preuves quant à la validité empirique des hypothèses nécessaires à une modélisation QALYs cohérente avec la théorie de l'utilité multi-attribut, en revanche, elle montre son intérêt pour modéliser des fonctions d'utilité pour des états de santé chroniques. Ainsi les QALYs peuvent constituer une aide en fournissant une mesure pouvant rendre les décisions de santé plus explicites mais ils ne permettent pas bien sûr de rendre complètement compte du caractère multidimensionnel de la santé.

# RESULTATS DES TRAVAUX DE RECHERCHE

Bien que le concept théorique d'utilité sous-tende qu'à chaque état de santé correspond une seule valeur d'utilité indépendamment de la façon de la mesurer, différents instruments peuvent donner différentes valeurs. Il existe très peu de données dans la polyarthrite débutante, population d'intérêt car candidate, en vertu de la notion de fenêtre thérapeutique, aux nouvelles biothérapies, traitements très coûteux. Actuellement, il n'y a pas de consensus sur le choix du questionnaire multi-attributs à utiliser pour le calcul des utilités. Or ce choix peut clairement modifier les résultats des études sur les nouveaux agents biologiques, les différentes stratégies thérapeutiques et leur rapport coût-utilité. L'objectif de cette thèse d'université est donc d'étudier les niveaux, les performances et les déterminants des 2 scores d'utilité indirecte (EQ-5D et SF-6D) dans la polyarthrite débutante afin de déterminer si un des deux instruments est plus pertinent. Tous les travaux ont été conduits dans la cohorte ESPOIR, cohorte multicentrique française de polyarthrites débutantes.

- Le premier travail a porté sur l'étude des propriétés psychométriques de l'EQ-5D et du SF-6D à partir des données de la cohorte ESPOIR dans une approche transversale. Ont été évaluées la concordance entre les deux scores, la validité de construction et la capacité discriminante.

- Le deuxième travail abordait l'étude longitudinale de l'EQ-5D et du SF-6D dans la cohorte ESPOIR. L'objectif de ce travail était de décrire et de comparer la sensibilité au changement de l'EQ-5D et du SF-6D.

- L'EQ-5D autorise des utilités négatives donc des états de santé « pires que la mort » ce que ne permet pas le SF-6D. L'objectif de cette 3° étude était d'identifier les patients avec EQ-5D négatif dans la cohorte ESPOIR et de décrire leurs caractéristiques. Le but était de déterminer quels aspects spécifiques (socio-démographiques, activité, incapacité, psychologique…) de la polyarthrite débutante étaient associés aux états de santé « pires que la mort ».

- La 4° étude portait sur le sens de l'utilité et de la différence entre EQ-5D et SF-6D. L'objectif de cette étude transversale réalisée à l'inclusion était de : 1) déterminer quelles variables étaient associées avec l'utilité EQ-5D et l'utilité SF-6D dans la polyarthrite débutante 2) déterminer quelles variables étaient associées à une différence substantielle d'utilité entre l'EQ-5D et le SF-6D.

- Dans la 5° étude, nous nous sommes ensuite intéressés à l'évolution au cours du temps des déterminants de l'utilité EQ-5D, SF-6D et de la différence entre les deux mesures d'utilité.

- Enfin, le 6° et dernier travail longitudinal a porté sur la finalité des utilités, au travers de l'étude des changements d'utilité EQ-5D et SF-6D au cours du temps. Les gains de QALYs avec l'EQ-5D et le SF-6D sous différentes stratégies thérapeutiques ont également été calculés et comparés.

## I – PROPRIETES PSYCHOMETRIQUES DE L'EQ-5D ET DU SF-6D EN TRANSVERSAL

Mon premier travail a porté sur l'étude des propriétés psychométriques de l'EQ-5D et du SF-6D à partir de la cohorte ESPOIR dans une approche transversale.

L'EQ-5D et le SF-6D ont été recueillis prospectivement pendant 2 ans dans une cohorte de 813 patients souffrant de polyarthrite débutante. Ont été évaluées la concordance entre les deux scores (coefficients de corrélations intra-classe [ICC], graphes de Bland-Altman), la validité de construction (corrélation avec différents marqueurs de l'état de santé d'un patient souffrant de polyarthrite) et la capacité discriminante (efficacité relative).

- A J0, le score moyen d'utilité avec l'EQ-5D était de 0,52 ± 0,31 (étendue -0,59 à 1) et de 0,58 ± 0,11 (étendue 0,30 à 0,92) pour le SF-6D. La distribution de l'EQ-5D était bimodale et celle du SF-6D, normale. Aucun effet plancher ni plafond n'était mis en évidence.

- La concordance entre les 2 instruments était faible chez les patients aux extrémités de l'échelle surtout en cas de forte incapacité ou de maladie active. Quand l'état de santé était mauvais, l'utilité était systématiquement plus basse avec l'EQ-5D qu'avec le SF-6D. A l'inverse, pour les patients en bonne santé, l'utilité était systématiquement plus haute avec l'EQ-5D qu'avec le SF-6D. Dans les essais cliniques et les analyses coût-efficacité où les patients ont des polyarthrites très actives, le gain en utilité sera plus important avec l'EQ-5D et le rapport coût –utilité sera plus favorable.

- La concordance était faible à J0: ICC=0,42, augmentant à 0,53 à 6 mois et 0,57 à 1 et 2 ans. Cette tendance à l'amélioration de la concordance au fur et à mesure du suivi reflète essentiellement l'amélioration de l'état de santé des patients avec peu de patients qui vont très mal à partir de 6 mois. Par contre, on observe également une meilleure concordance chez les patients avec une maladie peu active (avec une variabilité qui diminue nettement d'ailleurs) : cela pourrait indiquer une meilleure compréhension du questionnaire ou une adaptation au diagnostic récent.

- Les corrélations entre le Health Assessment Questionnaire (HAQ, incapacité fonctionnelle), le DAS 28 (activité de la PR), le score physique du SF36 (fonction) et les 2 instruments d'utilité étaient modérées à bonnes, stables dans le temps et similaires. Par contre, il existait une nette amélioration des corrélations avec le score mental du SF36 et la douleur à 6 mois

(initialement médiocre à modérée) puis stabilisation avec une meilleure corrélation pour le SF-6D.

- L'EQ-5D et le SF-6D différenciaient de manière identique les patients en faible activité mais le SF-6D discriminait mieux les patients avec une maladie très active ou une incapacité fonctionnelle (HAQ>1).

L'EQ-5D et le SF-6D ne sont donc pas interchangeables ; de ces résultats, il n'est pas possible d'indiquer une supériorité de l'un des outils par rapport à l'autre ***(J Rheum 2011, AO, 1er auteur, IF=3,854).***

**Article 1**

Comparison of the EQ-5D and the Short Form 6D utility measures in French patients with early arthritis: Results from the ESPOIR cohort

**Gaujoux-Viala C**, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, Fautrel B

J Rheumatol. 2011 Aug;38(8):1576-84.

## II – SENSIBILITE AU CHANGEMENT DE L'EQ-5D ET DU SF-6D

Mon deuxième travail a porté sur l'étude longitudinale de l'EQ-5D et du SF-6D dans la cohorte ESPOIR. En effet, la notion d'instauration rapide d'un traitement efficace et agressif dans la polyarthrite débutante a renforcé la nécessité de disposer de mesures d'utilité validées afin d'optimiser l'évaluation médico-économique de ces traitements. Mais les données sur leur sensibilité au changement sont rares et discordantes chez les patients souffrant de rhumatisme inflammatoire. L'objectif de ce travail était de décrire et de comparer la sensibilité au changement de l'EQ-5D et du SF-6D dans une cohorte multicentrique de patients souffrant de polyarthrite débutante.

La sensibilité au changement a été évaluée en calculant la réponse moyenne standardisée (RMS) à 6 et 12 mois dans des sous-groupes de patients caractérisés par leur évolution (amélioration ou détérioration du DAS28 > 0). La sensibilité au changement du HAQ a également été calculée afin de servir de point de comparaison.

L'EQ-5D présentait une moyenne de changement et une variance plus importante quelque soit la direction du changement. Les RMS de l'EQ-5D et du SF-6D étaient plus importantes en cas d'amélioration du DAS28 à 12 mois qu'en cas de détérioration.
Le SF-6D avait une plus grande sensibilité au changement en cas d'amélioration, comparable à celle du HAQ: RMS SF-6D = 0,83 versus 0,55 pour l'EQ-5D à 12 mois.
L'EQ-5D était plus sensible au changement en cas de détérioration avec une RMS proche de celle du HAQ : RMS EQ-5D = 0,20 versus 0,08 pour le SF-6D à 12 mois.
Les résultats étaient similaires à 6 mois. Si l'on sélectionnait les patients ayant les mêmes caractéristiques que ceux recrutés dans les essais cliniques testant l'efficacité des biothérapies, les conclusions étaient identiques.

Ce travail montre que la comparaison de la sensibilité au changement de l'EQ-5D et du SF-6D diffère selon la direction du changement. Le SF-6D est plus sensible à l'amélioration que l'EQ-5D. Par contre, l'EQ-5D est plus sensible à la détérioration.
La moyenne de changement en valeur absolue est plus importante pour l'EQ-5D que pour le SF-6D, ce qui a des implications directes sur le calcul des rapports coût-utilité qui seront donc plus favorables avec l'EQ-5D.

Le SF-6D est plus conservatoire. Enfin, l'estimation du gain d'utilité avec le SF-6D est plus précise car sa variabilité est toujours plus faible qu'avec l'EQ-5D.

Sur ces premiers arguments, l'utilisation du SF-6D paraît plus appropriée pour l'évaluation médico-économique des nouveaux traitements chez les patients souffrant d'arthrite débutante ***(ARD, AO, 1er auteur, IF=9,082)***.

**Article 2**

Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort.

**Gaujoux-Viala C**, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, Fautrel B.

Ann Rheum Dis. 2012 Sep;71(9):1478-83.

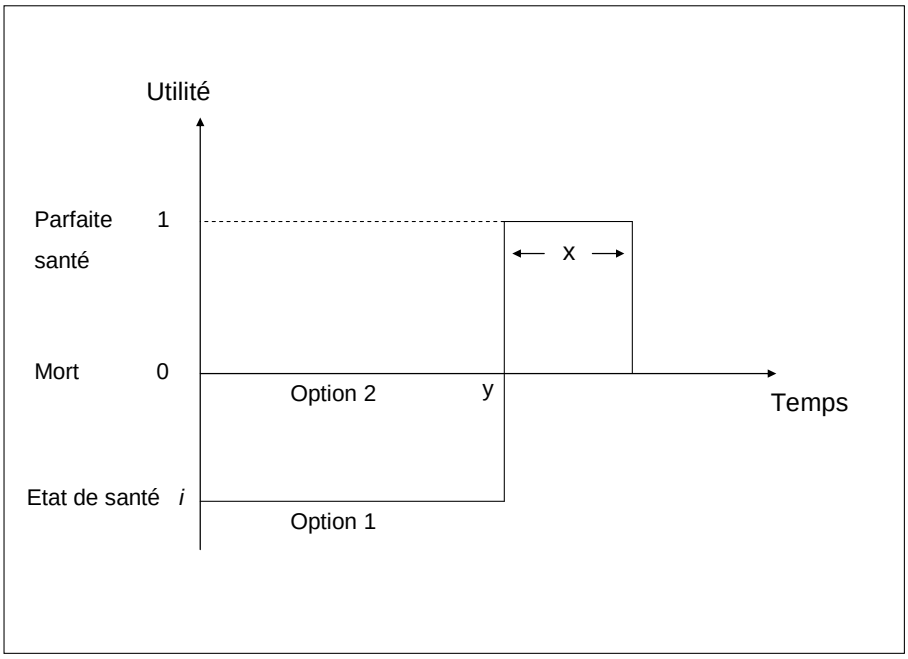
## III – LES PATIENTS AVEC UNE UTILITE NEGATIVE SUR L'EQ-5D, UN ETAT DE SANTE « PIRE QUE LA MORT

L'EQ-5D et le SF-6D donnent des résultats différents. Notamment l'EQ-5D, valorisé par arbitrage temporel, autorise des utilités négatives donc des états de santé « pires que la mort » ce que ne permet pas le SF-6D. Dans la procédure de l'arbitrage temporel pour les états « pires que la mort », le répondant est toujours confronté à deux options :

- La première (option 1) combine un état de santé i pendant y années suivi de la pleine santé pendant x années (avec x + y = t), ces t ans sont suivis du décès.
- L'option 2 correspond à un décès certain.

Le temps x varie jusqu'à ce que le répondant soit indifférent entre les deux options. A ce point, l'utilité de l'état de santé i est alors défini par x /(t - x) (Figure 17).

**Figure 17 :** Principe de l'arbitrage temporel pour les états pires que la mort (avec t=10 ans) (68)




L'objectif de cette étude était d'identifier les patients avec une utilité EQ-5D négative dans la cohorte de polyarthrites débutantes ESPOIR et de décrire leurs caractéristiques. Le but était de déterminer quels aspects spécifiques (socio-démographiques, activité, incapacité, psychologique…) étaient associés aux états de santé « pires que la mort ».

L'EQ-5D a été recueilli de manière prospective pendant 2 ans dans une cohorte de 813 patients souffrant de polyarthrite débutante (ESPOIR). Ont également été évalués le DAS28, le HAQ, le SF36 et des paramètres biologiques. Les patients avec un EQ-5D <0 ont été identifiés ; leur état de santé et leurs caractéristiques ont été analysés. Une régression logistique multivariée a été utilisée pour déterminer quels aspects spécifiques (socio-démographiques, activité, incapacité, psychologique…) étaient associés aux états de santé « pires que la mort ».

A J0, 90 (11%) des patients présentaient un état de santé « pire que la mort ». L'EQ-5D médian était de -0,052 (étendue -0,530 à -0,011). Ils présentaient presque tous un niveau maximum de douleur/inconfort avec en plus un niveau au moins modéré dans le domaine anxiété/dépression.
En analyse bivariée, les patients avec un EQ-5D négatif avaient moins souvent fait d'études supérieures, présentaient une maladie plus active, un HAQ plus élevé (incapacité fonctionnelle), plus de douleur, plus de fatigue, un score résumé mental du SF36 plus bas et un taux de CRP plus élevé que les patients avec un EQ-5D ≥ 0 (tous les p < 0,0001).
Dans le modèle de régression logistique, seules l'augmentation du HAQ et une diminution du score résumé mentaI du SF36 étaient associées avec un EQ-5D négatif.
Les résultats étaient similaires à 6, 12 et 24 mois avec 25, 23 et 24 patients dans un état de santé « pire que la mort».

Dans la polyarthrite débutante, la douleur/l'inconfort est le domaine clé conduisant à un EQ-5D < 0 correspondant à un état « pire que la mort », évalué au niveau le plus élevé. L'analyse mutivariée révèle que seuls le HAQ et le score résumé mental du SF36 sont toujours indépendamment associés aux états pires que la mort. Cette étude met en évidence des besoins spécifiques chez les patients souffrant de polyarthrite débutante nécessitant une meilleure prise en charge de la douleur et de l'anxiété/dépression.
***(Rheumatology, AO, 1° auteur, IF=4,749)***

**Article 3**

Who are the patients with early arthritis with “worse than death” scores on the EQ-5D? Results from the ESPOIR cohort

**Gaujoux-Viala C**, Fautrel B, Guillemin F, Flipo RM, Bourgeois P, Rat AC

Rheumatology 2012 [Epub ahead of print]

## IV – DETERMINANTS DE L'UTILITE EQ-5D, SF-6D ET DE LA DIFFERENCE ENTRE LES DEUX

Nous nous sommes ensuite intéressés à la question du sens : qu'est-ce qui constitue les mesures d'utilité EQ-5D et SF-6D ? Et qu'est-ce qui explique leur différence à un temps donné?

L'objectif était de : 1) déterminer quelles variables étaient associées avec l'utilité mesurée par l'EQ-5D et le SF-6D dans la polyarthrite débutante 2) déterminer quelles variables étaient associées avec une différence substantielle d'utilité entre les deux mesures.

L'EQ-5D et le SF-6D ont été recueillis chez 813 patients souffrant de polyarthrite débutante. Les déterminants de l'utilité EQ-5D et SF-6D à l'inclusion ont été analysés par régression linéaire multivariée. En raison de la distribution bimodale de l'utilité EQ-5D avec une séparation à 0,3, une régression logistique multivariée a également été utilisée pour déterminer quels aspects étaient associés à un score d'utilité EQ-5D<0,3. Des modèles multivariés de régression linéaire ont été utilisés pour déterminer quels aspects spécifiques de la polyarthrite débutante étaient associés à une différence d'utilité chez les patients présentant une différence qualifiée de substantielle car supérieure à la différence minimale importante (MID) : $|$SF-6D – EQ-5D$|$>0,074.

Les patients ont été classés en 3 catégories :

- SF-6D substantiellement supérieur à l'EQ-5D ($|$SF-6D – EQ-5D$|$>0.074 et SF-6D>EQ-5D),
- EQ-5D substantiellement supérieur au SF-6D ($|$SF-6D – EQ-5D$|$>0.074 et EQ-5D>SF-6D)
- et les autres correspondant à l'absence de différence substantielle entre les deux mesures d'utilité (catégorie de référence).

Une régression polynomiale a été utilisée avec cette variable dépendante à 3 catégories.

À J0, la valeur moyenne de l'utilité EQ-5D était de 0.52 ± 0.31, médiane 0.66 (étendue -0,59 à 1). La valeur moyenne de l'utilité SF-6D était de 0,58 ± 0,11, médiane 0,58 (étendue 0,30 à 0,92).

Dans le modèle de régression linéaire multivariée, une plus grande incapacité fonctionnelle (HAQ), une altération de l'état mental (AIMS2-SF) et une douleur au repos plus élevée étaient associées à une utilité EQ-5D plus basse. Le HAQ expliquait 41,4% de la variance et le score mental, 7,9%. Ces résultats ont été confirmés par la régression logistique : une plus

grande incapacité fonctionnelle, une altération de l'état mental, une douleur au repos plus élevée et un niveau éducatif plus bas (absence d'études supérieures) étaient associées à une utilité EQ-5D <0,3.

Une plus grande incapacité fonctionnelle, une altération de l'état mental, une douleur au repos plus élevée, une évaluation globale de l'activité par le patient plus importante, une plus grande fatigue, et le fait d'avoir un niveau éducatif plus bas étaient associés à une utilité SF-6D plus faible. Le HAQ expliquait 49,3% de la variance et la composante mentale, 14,5%.

La différence entre l'utilité SF-6D et l'EQ-5D était en moyenne de 0,064 [étendue -0,421; 0,941]. La majorité des patients (66,5%) présentaient une différence substantielle entre les deux mesures d'utilité (│SF-6D – EQ-5D│>0,074, différence minimale importante). Chez ces patients, une différence plus élevée était associée à une plus grande incapacité fonctionnelle, un moins bon état mental et une CRP plus élevée. Le HAQ expliquait 43,3% de la variance.

Dans le modèle de régression polynomiale, l'incapacité fonctionnelle (HAQ), un moins bon état mental et une douleur plus élevée étaient associés avec la probabilité d'une différence substantielle d'utilité. L'influence du HAQ était bidirectionnelle : une incapacité fonctionnelle plus sévère (HAQ plus haut) était associée à la probabilité que l'utilité SF-6D soit substantiellement supérieure à l'utilité EQ-5D et une incapacité fonctionnelle moins sévère (HAQ plus bas) était associée à la probabilité que l'utilité EQ-5D soit substantiellement supérieure à l'utilité SF-6D. A l'inverse, un moins bon état mental et une douleur au repos plus élevée étaient associés à la probabilité d'avoir une différence substantielle entre les deux mesures d'utilité que ce soit l'utilité SF-6D ou l'utilité EQ-5D la plus élevée.

Une analyse de sensibilité a été effectuée avec d'autres valeurs de différence minimale importante à 0,06 et 0,129. Les résultats étaient similaires sauf à 0,129 où 2 variables supplémentaires devenaient significatives : une activité plus élevée (DAS28 plus grand) était associée à la probabilité que l'utilité EQ-5D soit substantiellement supérieure à l'utilité SF-6D et un niveau d'éducation plus bas (absence d'études supérieures) était associé à la probabilité que l'utilité SF-6D soit substantiellement supérieure à l'utilité EQ-5D.

Une deuxième analyse de sensibilité a été conduite en utilisant le score physique du AIMS2-SF à la place du HAQ confirmant que les différences substantielles d'utilité entre les deux instruments d'utilité étaient déterminées essentiellement par la fonction.

En conclusion, ce travail a montré que la fonction, et plus particulièrement le HAQ, est le déterminant majeur conduisant à des différences substantielles d'utilité entre l'EQ-5D et le SF-6D. L'EQ-5D et le SF-6D ne sont pas interchangeables surtout chez les patients avec une incapacité fonctionnelle importante (c'est à dire la grande majorité des patients rhumatologiques) et les résultats des études coût-utilité utilisant différents instruments ne doivent pas être comparés.

**Article 4**

The impact of functional disability on EQ-5D and SF-6D utility discrepancies in early arthritis: Results from the ESPOIR cohort

**Gaujoux-Viala C**, Fautrel B, Hosseini K, Guillemin F, Flipo RM, Rat AC.

Soumis

## V – EVOLUTION DES DETERMINANTS DE L'UTILITE EQ-5D, SF-6D ET DE LA DIFFERENCE ENTRE LES DEUX AU COURS DU TEMPS

Nous avons ensuite voulu savoir comment évoluaient les déterminants de l'utilité EQ-5D, SF-6D et les différences entre les deux à différents moments afin de savoir si l'on retrouvait toujours les mêmes déterminants et si les associations étaient stables au cours du temps.

L'EQ-5D et le SF-6D ont été recueillis chez 813 patients souffrant de polyarthrite débutante. Les déterminants de l'utilité EQ-5D et de l'utilité SF-6D ont été analysés par régressions linéaires multivariées répétées à l'inclusion, à 6 mois, 1, 2 et 3 ans. Des modèles multivariés de régression linéaire ont également été utilisés pour déterminer quels aspects spécifiques de la polyarthrite débutante étaient associés à une différence d'utilité chez les patients présentant une différence qualifiée de substantielle car supérieure à la différence minimale importante (MID) : $|\text{SF-6D} - \text{EQ-5D}| > 0.074$.

A J0, une plus grande incapacité fonctionnelle (HAQ), une altération de l'état mental (AIMS2-SF) et une douleur au repos plus élevée étaient associées à une utilité EQ-5D plus basse. Le HAQ expliquait 41,4% de la variance et le score mental, 7,9%. Les résultats étaient quasi-similaires aux autres temps excepté à 1 an où il y avait une inversion concernant les 2 déterminants principaux: l'état mental devenant prépondérant, expliquant 41,7 % de la variance et le HAQ 11,4 %.

A J0, une utilité SF-6D plus faible était associée à une plus grande incapacité fonctionnelle, une altération de l'état mental, une douleur au repos plus importante, une évaluation globale de l'activité par le patient plus grande, une plus grande fatigue, et le fait d'avoir un niveau éducatif plus bas. Le HAQ expliquait 49,3% de la variance et le score mental, 14,5%.

Les déterminants de l'utilité SF-6D restaient globalement les mêmes au cours du temps. Cependant dès 6 mois, il y avait une inversion concernant les 2 déterminants principaux: l'état mental devenant prépondérant, expliquant 55 à 57 % de la variance et le HAQ entre 7 et 13%.

Les différences substantielles entre l'utilité SF-6D et l'utilité EQ-5D étaient essentiellement déterminées par la fonction, exceptée à 1 an où elle reposait sur la douleur et l'état mental.

En conclusion, la fonction et l'état mental sont les déterminants principaux de l'utilité EQ-5D et de l'utilité SF-6D au cours des 3 premières années de la maladie. Une incapacité fonctionnelle importante et un mauvais score mental sont les éléments clés conduisant à une différence substantielle entre les deux mesures d'utilité.

La variabilité des déterminants de l'utilité SF-6D et EQ-5D au cours du temps, les difficultés de modélisation linéaire de l'EQ-5D posent question quant à l'utilisation du mapping pour estimer l'utilité à partir d'instruments de mesure de l'état de santé notamment à partir du HAQ.

L'évaluation médico-économique des différents traitements ne devrait pas reposer sur des mesures d'utilité EQ-5D ou SF-6D estimées à partir du HAQ, qui ignore l'impact mental de la maladie. Les futures études cliniques devraient inclure au moins une mesure d'utilité indirecte.

**Article 5**

The determinants of the SF-6D and EQ-5D utility measures vary over time in early arthritis: Results from the ESPOIR cohort

**Gaujoux-Viala C**, Rat AC, Hosseini K, Guillemin F, Flipo RM, Fautrel B

## VI – LES CHANGEMENTS D'UTILITE AVEC L'EQ-5D ET LE SF-6D

Ce travail longitudinal porte sur la finalité des utilités, au travers de l'étude des changements d'utilité EQ-5D et SF-6D. Les gains de QALYs avec l'EQ-5D et le SF-6D à 1 an dans l'ensemble de la population et sous différentes stratégies thérapeutiques ont également été calculés et comparés. L'article est en cours de rédaction.

### VI.1– Evolution de l'état de santé des patients souffrant de polyarthrite débutante

Au cours des six premiers mois, il y a une nette amélioration de l'état de santé moyen des patients, sur tous les plans (Table 1). Cette amélioration se poursuit jusqu'à 18 mois mais de manière moins marquée puis se stabilise (Figure 1).

En fait, si l'on regarde dans le détail, à 6 mois, sur le plan de l'activité de la PR (DAS 28), 622 patients s'améliorent et 97 s'aggravent. A 1 an, 646 patients voient l'activité de leur maladie diminuer et 61 augmenter.

**Table 1:** Characteristics of patients included in ESPOIR cohort at baseline, 6 months and 3 years

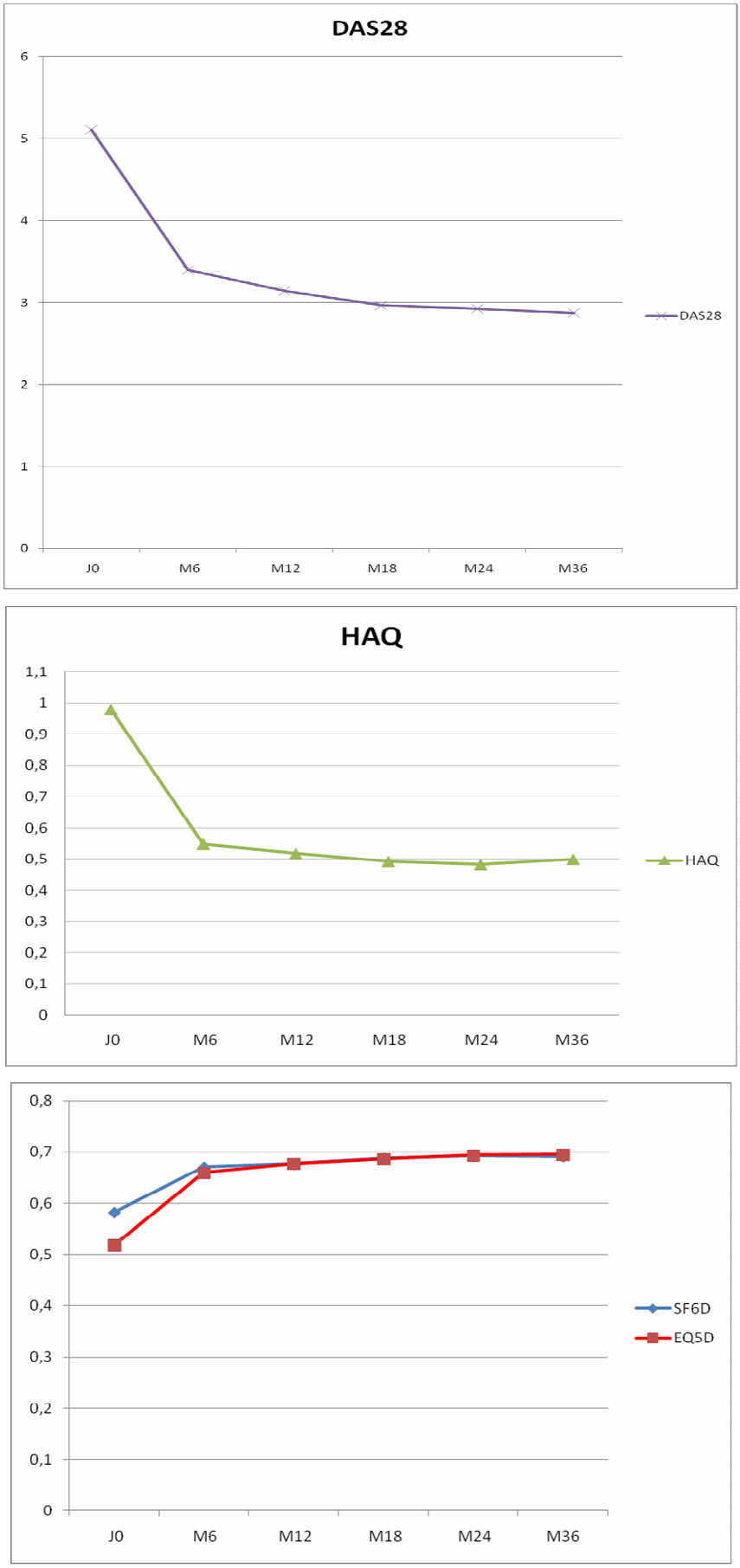
| Characteristic | **All patients at baseline (n=813)** Mean ± SD | **All patients at 6 months (n=757)** Mean ± SD | **All patients at 3 years (n=618)** Mean ± SD |
|---|---|---|---|
| Age (years) | 48.1 ± 12.6 | 48.8 ± 12.5 | 51.7 ± 12.0 |
| Female sex, N (%) | 624 (76.7) | 578 (76) | 475 (76.9) |
| >12 years of education, N (%) | 255 (31.4) | 244 (32.2) | 200 (32.3) |
| Married or living together, N (%) | 594 (73.1) | 573 (75.7) | 464(75.1) |
| DAS28 (0.49 -9.07 : with 9.07 high activity) | 5.11 ± 1.31 | 3.39 ± 1.40 | 2.87 ± 1.36 |
| HAQ (0-3: with 3 high disability) | 0.979 ± 0.684 | 0.548 ± 0.566 | 0.499 ± 0.580 |
| AIMS2-SF-mental score (0-100, with 100 the best mental status) | 58.3 ± 19.9 | 66.2 ± 20.4 | 69.3 ± 20.2 |
| Patient Global Assessment (0-100, with 100 the worst health) | 59.8 ± 25.6 | 34.7 ± 26.8 | 29.0 ± 26.2 |
| Pain at rest (0-100, with 100 high pain) | 37.2 ± 27.6 | 23.3 ± 24.0 | 19.6 ± 23.0 |
| Fatigue (0-100, with 100 the worst fatigue) | 47.8 ± 28.2 | 39.0 ± 28.4 | 33.6 ± 28.7 |
| Erythrocyte sedimentation rate | 29.4 ± 24.6 | 16.2 ± 14 | 14.9 ± 13.3 |
| C-reactive protein level * | 20.3 ± 32.4 | 8.8 ± 19.9 | 6.8 ± 11.3 |
| Rheumatoid factor*, N (%) | 372 (45.8) | 347 (45.8) | 300 (48.5) |
| Anti-CCP2 antibodies*, N (%) | 315 (38.8) | 296 (39) | 268 (43.4) |
| V H Sharp score $ | 5.7 ± 7.70 | NA | 13.9 ± 14.7 |
| SF-6D score | 0.58 ± 0.11 | 0.67 ± 0.12 | 0.69 ± 0.13 |
| EQ-5D score | 0.52 ± 0.31 | 0.66 ± 0.23 | 0.70 ± 0.24 |

NA: not available DAS28: disease activity score in 28 joints; HAQ: Health Assessment Questionnaire; AIMS2-SF: Arthritis Impact Measurement Scale 2–Short Form

*Baseline C-reactive protein level (normal<10 mg/l); IgM and IgA rheumatoid factor (Elisa, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive > 50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

$ Van der Heijde modified Sharp score.

**Figure 1:** Evolution of early arthritis patients health states over 3 years



### VI.2– Description des changements d'utilité

Les changements moyens d'utilité EQ-5D et SF-6D entre M0 et M6, M6 et M12, M12 et M24, M24 et M36 sont présentés dans le tableau 2.

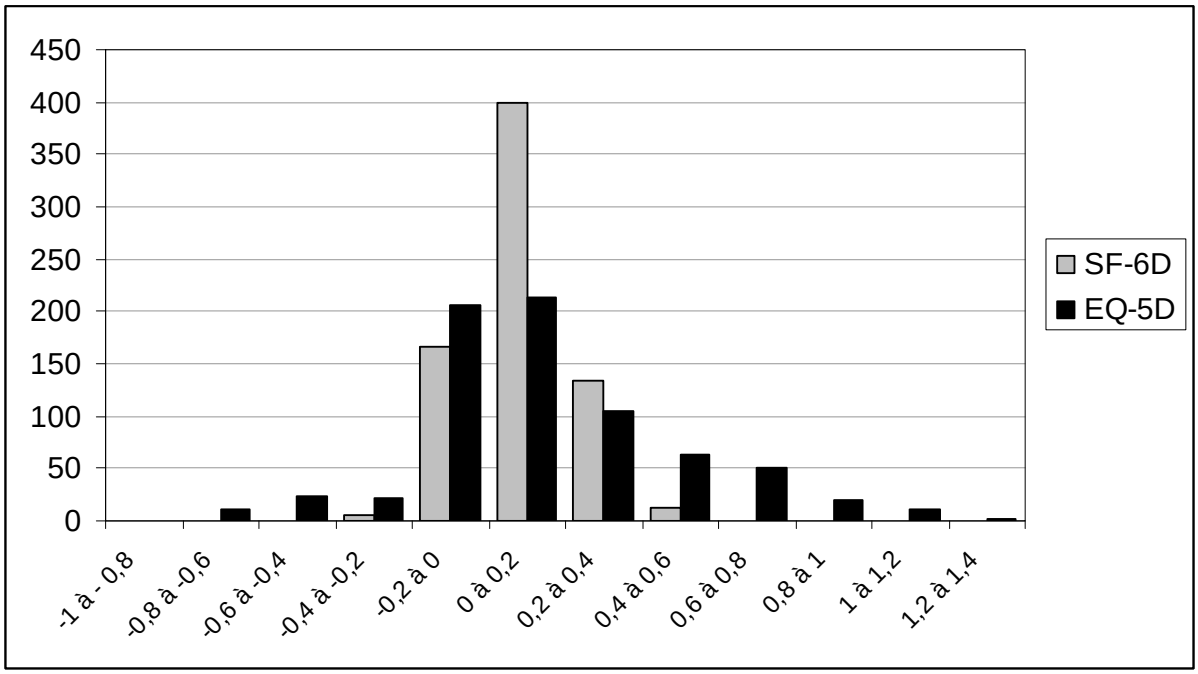
**Table 2:** Mean change of the EQ-5D and SF-6D between baseline and 6 months, 6 and 12 months, 12 and 24 months and, 24 and 36 months

| *Outcomes* *Mean ± DS* | ***M0-M6*** | ***M6-M12*** | ***M12-M24*** | ***M24-M36*** |
|---|---|---|---|---|
| SF-6D change (positive : improvement) | 0,088 ± 0,122 | 0,0085 ± 0,104 | 0,0142 ± 0,110 | -0,0014 ± 0,105 |
| EQ-5D change (positive : improvement) | 0,137 ± 0,300 | 0,0189 ± 0,223 | 0,0151 ± 0,236 | 0,0089 ± 0,227 |
| HAQ change (negative : improvement) | -0.421 ± 0.642 | -0.0307 ± 0.397 | -0.0283 ± 0.444 | 0.009 ± 0.381 |
| DAS 28 change (negative : improvement) | -1.72 ± 1.49 | -0.24 ± 1.23 | -0.19 ± 1.25 | -0.06 ± 1.27 |
| AIMS2-SF-mental score change (positive : improvement) | 7.5 ± 18.8 | 0.39 ± 16.6 | 2.25 ± 16.2 | 0.28 ± 15.8 |
| Pain change (negative : improvement) | -13.5 ± 29.7 | -0.8 ± 25.2 | -2.2 ± 24.1 | -0.5 ± 23.5 |

Le changement moyen d'utilité et la variabilité sont toujours plus grands avec l'EQ-5D qu'avec le SF-6D. Nous avons comparé les changements d'utilité EQ-5D et SF-6D par intervalle avec un test de Student apparié. Les différences entre les changements d'utilité avec les deux scores sont significatives seulement pour le changement entre l'inclusion et 6 mois : différence moyenne de 0,05 ± 0,26 en faveur de l'EQ-5D ($p<0,0001$).

Alors que la distribution de l'utilité EQ-5D est bimodale et celle du SF-6D quasi-normale, les distributions des changements d'utilité EQ-5D et SF-6D sont proches de la normalité (Figure 2).

**Figure 2:** Frequency distribution of the change in EQ-5D and SF-6D utility scores at 1 year




## VI.3– Comparaison des QALYs obtenus à 1 an à partir de l'EQ-5D et du SF-6D

### VI.3.1– Dans l'ensemble de la population

Nous avons calculé et comparé les QALYs obtenus à 1 an à partir de l'EQ-5D et SF-6D avec un test de Student apparié. Le gain de QALYs est de 0,109 ± 0,214 pour l'utilité EQ-5D versus 0,067 ± 0,084 pour l'utilité SF-6D. La différence de 0,042 ± 0,185 QALYs en faveur de l'EQ-5D est significative ($p<0,0001$).

### VI.3.2– Chez les patients sous méthotrexate sans biothérapie pendant les 6 premiers mois

Il s'agit d'un exemple de stratégie thérapeutique. Un patient est considéré sous méthotrexate s'il a reçu au moins 3 mois de méthotrexate pendant les 6 premiers mois. Ont été exclus les patients sous biothérapies pendant les 6 premiers mois.

En tout, 313 patients ont reçu au moins 3 mois de méthotrexate. Les caractéristiques des patients sont décrites dans le tableau 4.

**Table 4:** Characteristics of patients in ESPOIR cohort included in this study at baseline

| **Characteristic** | **All patients (n=777)** Mean ± SD | **MTX (n=313)** Mean ± SD |
|---|---|---|
| Age (years) | 48.0±12.7 | 48.8±12.0 |
| Female sex, N (%) | 598 (77.0%) | 237 (75.7%) |
| >12 years of education, N (%) | 245 (31.5%) | 97 (31.0%) |
| Employed, N (%) | 492 (63.3%) | 207 (66.1%) |
| Smoking | 370 (47.6%) | 155 (49.5%) |
| DAS28 | 5.06±1.30 | 5.39±1.22 |
| HAQ | 0.96±0.67 | 1.11±0.68 |
| Symptom duration (months) | 4.7±5.7 | 4.9±5.9 |
| C-reactive protein level* | 19.8±32.0 | 24.9±37.4 |
| Rheumatoid factor, N (%)* | 342 (44.1%) | 171 (54.6%) |
| Anti-CCP2 antibodies*, N (%) | 287 (37%) | 156 (49.8%) |
| Erosion present | 403 (57.7%) | 191 (64.5%) |
| Use of corticosteroids # | 168 (26.0%) | 97 (35.0%) |
| SF-6D score | 0.58 ± 0.11 | 0.57 ± 0.11 |
| EQ-5D score | 0.52 ± 0.30 | 0.50 ± 0.31 |

*Baseline C-reactive protein level (normal<10 mg/l); IgM and IgA rheumatoid factor (Elisa, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive >50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

# Use of corticosteroids: at least 7.5 mg/day equivalent prednisone for more than 3 months in the first year.

La dose moyenne de méthotrexate était de 12,7±3,8 mg/semaine. La voie d'administration était orale dans la grande majorité des cas (97,4% des patients).

Nous avons calculé et comparé les QALYs obtenus à 1 an à partir de l'EQ-5D et SF-6D avec un test de Student apparié. Le gain de QALYs est de 0,118 ± 0,213 pour l'utilité EQ-5D versus 0,071 ± 0,082 pour l'utilité SF-6D. La différence de 0,050 ± 0,190 QALYs en faveur de l'EQ-5D est significative (p<0,0001).

Ce travail montre que les changements d'utilité SF-6D et EQ-5D reflètent bien l'évolution de l'état de santé des patients souffrant de polyarthrite débutante. L'état de santé moyen s'améliore à 6 mois puis atteint un plateau à 12 mois. Les changements d'utilité sont toujours plus importants avec l'EQ-5D ainsi que la variabilité de ces changements. Par contre, la différence n'est significative qu'entre 0 et 6 mois. Le gain de QALYs à 1 an est quasiment doublé avec l'EQ-5D par rapport au SF-6D pour une même stratégie thérapeutique, ce qui fournira des rapports coût-utilité plus favorables. Cependant la variabilité qui caractérise ce gain de QALYs est-elle aussi doublée, ce qui fait que le rapport coût-utilité sera entouré de plus d'imprécision.

# SYNTHESE

L'évaluation médico-économique est amenée à prendre une place de plus en plus importante dans l'évaluation des soins médicaux et des nouvelles thérapeutiques en raison de dépenses de santé en constante augmentation couplées à des ressources limitées. L'évaluation d'un gain de santé s'intéresse non seulement à l'état de santé des patients mais aussi à leur perception de cet état de santé et de son impact sur leur qualité de vie. L'utilisation d'auto-questionnaires multi-attributs permet de définir un ensemble d'états auxquels sont associés des niveaux d'utilité, estimés en population générale, qui tiennent compte des préférences de la collectivité vis-à-vis des différents états de santé. Actuellement, il n'y a pas de consensus sur le choix du questionnaire multi-attributs à utiliser pour le calcul des utilités et il existe très peu de données dans la polyarthrite débutante, dont le traitement a été révolutionné ces dernières années par l'arrivée des biothérapies très efficaces mais très coûteuses. L'objectif de cette thèse d'université est donc d'étudier les niveaux, les performances et les déterminants des 2 scores d'utilité indirecte (EQ-5D et SF-6D) dans la polyarthrite débutante afin de déterminer si un des deux instruments est plus pertinent.

A l'issue de ce programme de recherches, plusieurs enseignements peuvent être tirés.

- Dans une cohorte française de polyarthrites débutantes, la concordance entre les 2 instruments était faible chez les patients aux extrémités de l'échelle surtout en cas de forte incapacité ou de maladie active. Quand l'état de santé était mauvais, l'utilité était systématiquement plus basse avec l'EQ-5D qu'avec le SF-6D. A l'inverse, pour les patients en bonne santé, l'utilité était systématiquement plus haute avec l'EQ-5D qu'avec le SF-6D. Dans les essais cliniques et les analyses coût-efficacité où les patients ont initialement des polyarthrites très actives et bénéficient d'un traitement efficace, le gain en utilité sera plus important avec l'EQ-5D et le rapport coût –utilité sera plus favorable.

On pourrait penser que ce manque de concordance est la conséquence directe des différences d'échelles mais la réalisation d'analyse de sensibilité après standardisation des deux scores (centrés-réduits) ne montre qu'une amélioration modérée de la concordance.

La concordance était faible à J0 augmentant à partir 6 mois pour atteindre un plateau à 1 et 2 ans. Cette tendance à l'amélioration de la concordance au fur et à mesure du suivi pourrait s'expliquer par l'amélioration de l'état de santé des patients et en partie, indiquer une sorte de familiarisation avec l'instrument. Cela pourrait également résulter du phénomène d'adaptation des patients à leur maladie (*response shift)* (84).

L'EQ-5D et le SF-6D ne sont donc pas interchangeables.

1. Différences de distribution :

Dans une cohorte française de polyarthrites débutantes, la distribution des scores d'utilités EQ-5D était bimodale et celle des scores d'utilités SF-6D, normale. Cette distribution en deux parties de l'utilité EQ-5D a été retrouvée dans de nombreuses pathologies (89–91) et pose problème pour l'application des outils statistiques qui suppose, pour la plupart, une « quasi-normalité ». Par contre, aucun effet plancher ni plafond n'était mis en évidence.

2. Différences de propriétés psychométriques :

Pour qu'un instrument d'évaluation soit adapté, il doit posséder un certain nombre de propriétés psychométriques, définies par le « filtre OMERACT » (92) qui a été mis au point par un groupe de méthodologistes (OMERACT, Outcome Measures in Rheumatology) et qui définit ces diverses propriétés métrologiques, appelées *truth, feasibility, et discriminant capacity* :

- Truth (validité) implique que l'instrument mesure ce qu'il est sensé mesuré (93). Il s'agit de la qualité métrologique la plus importante. La validité de corrélation (concurrent validity) ou de critère (criterion validity) est évaluée par la corrélation entre le nouvel instrument et un outil qui a déjà fait ses preuves. On parle de validité de critère s'il s'agit de mesurer le critère contre un critère extérieur dit de référence ou 'gold standard'.

- Feasability (faisabilité) : la faisabilité fait référence à l'acceptabilité et englobe le temps nécessaire à l'examen, des contraintes pour le patient, du coût et /ou de la disponibilité des appareils de mesure…Il s'agit d'une propriété métrologique « à part » car elle ne se mesure pas avec des statistiques, sauf pour le temps de passation et éventuellement le taux de valeurs manquantes.

- Discriminant capacity (discrimination) englobe différentes propriétés :

La fiabilité (reliability) ou fidélité rend compte du degré d'erreur de mesure, ou capacité à donner des résultats comparables lorsque l'état est constant et que les situations expérimentales varient.

Un instrument de mesure est fiable s'il produit les mêmes scores lors de mesures effectuées à différents moments par le même examinateur (reproductibilité inter-temps) et lors de mesures effectuées au même moment par différents examinateurs (reproductibilité inter-juges). Dans le cas de questionnaires auto-remplis par un patient, il n'y a pas de reproductibilité inter-juges à évaluer.

On inclut également souvent dans la notion de reproductibilité, la précision d'une mesure.

La sensibilité au changement est l'aptitude d'une échelle à mettre en évidence un changement du statut clinique, entre 2 situations ou entre 2 moments donnés. La sensibilité au changement capte la différence entre des mesures faites à deux temps différents, par exemple sous l'effet d'un traitement, tandis que la capacité discriminante distingue 2 états (par exemple malade versus sain). La capacité discriminante et la sensibilité au changement sont particulièrement importantes pour un critère de réponse thérapeutique (94).

Une étude internationale COSMIN a récemment proposé un regroupement de ces propriétés psychométriques dans une nouvelle nomenclature mais les concepts restent identiques (95).

Nous avons donc étudié et comparé les différentes propriétés métrologiques des 2 instruments d'utilité indirecte EQ-5D et SF-6D :

- La faisabilité était intéressante avec très peu de valeurs manquantes : 1,2% à J0 et 1% à 2 ans pour l'utilité SF-6D, 0,6% à J0 et 0,3% à 2 ans pour l'utilité EQ-5D.

- La validité de construction était bonne pour l'utilité EQ-5D et SF-6D. Les corrélations entre l'incapacité fonctionnelle (HAQ, score physique du SF36), l'activité de la polyarthrite (DAS 28) et les 2 instruments d'utilité étaient modérées à bonnes, stables dans le temps et similaires. Par contre, les corrélations avec le score mental et la douleur qui étaient initialement médiocre à modérée s'amélioraient à 6 mois puis se stabilisaient avec une meilleure corrélation pour l'utilité SF-6D.

- La reproductibilité des deux instruments n'a pas pu être évaluée dans ce travail mais elles étaient bonnes dans la littérature (96) avec cependant un avantage en faveur du SF-6D (90,97).

- Concernant la capacité discriminante, l'EQ-5D et le SF-6D différenciaient de manière identique les patients en faible activité mais le SF-6D discriminait mieux les patients avec une maladie très active ou une incapacité fonctionnelle (HAQ>1).

- La comparaison de la sensibilité au changement de l'EQ-5D et du SF-6D différait selon la direction du changement. Le SF-6D avait une plus grande sensibilité au changement en cas d'amélioration. L'EQ-5D était plus sensible à la détérioration.
L'EQ-5D présentait une moyenne de changement en valeur absolue plus importante quelque soit la direction du changement, ce qui a des implications directes sur le calcul des rapports coût-utilité qui seront donc plus favorables avec l'EQ-5D. Le SF-6D est plus conservatoire. Enfin, l'estimation du gain d'utilité avec le SF-6D est plus précise car sa variabilité était toujours plus faible qu'avec l'EQ-5D.

Sur ces premiers arguments, l'utilisation du SF-6D paraît plus appropriée pour l'évaluation médico-économique des nouveaux traitements chez les patients souffrant d'arthrite débutante. Par contre, l'EQ-5D étant plus sensible à la détérioration, il pourrait trouver sa place dans les études de stratégie thérapeutique sur la décroissance et l'arrêt des traitements biologiques en cas de rémission persistante.

3. Différences de sens :
La comparaison des 2 outils d'utilité indirecte implique également la question du « sens » : qu'est-ce qui constitue ces mesures d'utilité ? Pourquoi sont-elles différentes ?

Tout d'abord, l'EQ-5D permet d'obtenir des utilités négatives donc des états de santé « pires que la mort » ce que ne permet pas le SF-6D. Cet état n'est pas rare dans la polyarthrite débutante puisqu'à l'inclusion, 11% des patients présentaient un état de santé « pire que la mort » avec un niveau maximum de douleur/inconfort et en plus, un niveau au moins modéré dans le domaine anxiété/dépression. Cette proportion diminuait dès 6 mois et restait stable au cours du temps autour de 3,3%.
En analyse multivariée, seuls l'incapacité fonctionnelle (HAQ) et un mauvais état mental (diminution du score résumé mental du SF36) étaient toujours indépendamment associées à un score d'utilité EQ-5D négatif et ce de manière constante pendant 2 ans.

Ces résultats étaient similaires à ceux d'une étude regroupant 2 essais au Royaume-Uni : l'un chez des patients souffrant de polyarthrites évoluant depuis moins de 11 semaines, et l'autre chez des patients avec PR évoluant depuis plus de 5 ans. Par contre, les résultats n'étaient pas différenciés pour ces deux populations (98). Sur 700 patients, 9% présentaient un score d'utilité négatif avec l'EQ-5D. Tous les patients présentaient un niveau maximum de douleur/inconfort et en plus, un niveau au moins modéré dans au moins 3 autres domaines. Adams et al. ont étudié 345 patients souffrant de PR établies, 17% présentaient un état de santé « pire que la mort » à J0 et 7% lors du suivi (99). Les patients souffrant de rhumatisme psoriasique présentaient moins souvent d'utilité EQ-5D négative (12% à J0 et 2% lors du suivi). Dans ces études, un score maximum de douleur était toujours fortement corrélé à ces états de santé "pires que la mort" (99).
La diminution de la proportion de patients avec une utilité EQ-5D négative au cours du temps pourrait traduire un effet bénéfique des traitements ou l'adaptation du patient à sa maladie (*response shift*) (84).

Il convient cependant de nuancer ce qualificatif d'état de santé « pire que la mort ». En effet, il s'agit de la perception d'un état de santé par la population générale. Macran et al ont identifié un petit groupe de sujets issus d'un échantillon représentatif de la population d'Angleterre qui confèrent une valeur relativement élevée à la mort par rapport à d'autres états de santé (100). De plus, nous ne savons pas si les patients seraient d'accord pour qualifier leur état de "pire que la mort" car leurs perceptions peuvent différer de celles de la population générale (101). Il serait intéressant d'approfondir ce point par le biais d'une enquête qualitative sur les perceptions des malades et non-malades.

En tous cas, l'impact de la PR débutante sur la douleur, la fonction et l'état psychologique révélé au travers de l'étude de ces « états de santé pires que la mort » met en lumière la nécessité d'une attention particulière sur la prise en charge de la douleur et de l'anxiété-dépression.

Nous avons démontré que l'EQ-5D et le SF-6D donnent des résultats différents et que la fonction joue un rôle particulier dans ces différences.
A J0, la fonction (HAQ) était le déterminant principal de l'utilité EQ-5D, SF-6D et de la différence entre les deux. L'incapacité fonctionnelle, un moins bon état mental et une augmentation de la douleur étaient associés à la probabilité d'une différence substantielle d'utilité.

L'influence du HAQ était de plus bidirectionnelle : une incapacité fonctionnelle augmentée (HAQ plus haut) était associée à la probabilité que l'utilité SF-6D soit substantiellement supérieure à l'utilité EQ-5D et une incapacité fonctionnelle diminuée (HAQ plus bas) était associée à la probabilité que l'utilité EQ-5D soit substantiellement supérieure à l'utilité SF-6D. A l'inverse, un moins bon état mental et une augmentation de la douleur au repos étaient associés à la probabilité d'avoir une différence substantielle entre les deux mesures d'utilité que ce soit l'utilité SF-6D ou EQ-5D la plus élevée.

Si initialement l'utilité SF-6D était constituée essentiellement de fonction (HAQ), dès 6 mois, elle était déterminée essentiellement par l'état mental. Que se passe-t-il entre ces deux temps ?

- Tout d'abord, il y a l'annonce du diagnostic et l'intégration par le patient du fait qu'il souffre d'une maladie chronique.
- Un élément parallèle est l'amélioration de l'état de santé des sujets. Initialement les patients avaient des douleurs importantes, une incapacité fonctionnelle marquée, un mauvais état mental. Dès 6 mois, la situation s'améliorait sur tous les plans.
- Cela peut-être dû à la prise en charge thérapeutique mais aussi au phénomène de *response shift*. Il s'agit d'un changement de la manière dont une personne auto-évalue sa qualité de vie :

- modification des mesures internes de l'individu (recalibration, en termes psychométriques)
- changement de la valeur accordée à un état de santé (importance relative des différents domaines constituants la qualité de vie)
- redéfinition de la qualité de vie (reconceptualisation) (102,103).

Ainsi Sprangers et al. ont voulu savoir s'il existait un changement des mesures internes de la fatigue chez les patients traités par radiothérapie pour un cancer (104). Ils ont comparé une évaluation rétrospective (après traitement) de l'état de fatigue initial avec l'évaluation effectivement réalisée avant traitement. Un phénomène de « response shift » a été retrouvé dans deux sous-groupes de patients :

- les patients qui voyaient leur niveau de fatigue diminuer avec le traitement, amplifiaient rétrospectivement leur fatigue initiale
- à l'inverse, les patients dont la fatigue avait augmenté minimisaient rétrospectivement leur niveau de fatigue avant traitement.

Ce phénomène de *response shift* ne peut que complexifier l'évaluation des états de santé. En effet, il a été mis en évidence que la valorisation des états de santé était dépendante de l'état de santé dans lequel se trouvait le répondant. Par exemple, Lenert et al. ont montré que les patients en mauvaise santé accordaient une valeur presque similaire aux états de santé

intermédiaires et aux états de santé quasi-normaux (105). A l'inverse, les patients en bonne santé conféraient à ces mêmes états de santé intermédiaires, une valeur proche cette fois-ci des mauvais états de santé. On peut cependant espérer qu'une partie de ce phénomène soit atténuée par l'utilisation d'utilités indirectes où le patient décrit son état de santé à partir d'un auto-questionnaire mais où les pondérations sont issues de la population générale.
Ces résultats laissent à penser que tant qu'il y a un niveau important d'incapacité fonctionnelle, c'est la fonction (HAQ) qui détermine l'utilité SF-6D mais qu'en dessous d'un certain seuil d'incapacité fonctionnelle, c'est le mental qui devient prépondérant.
Pour l'utilité EQ-5D, les déterminants étaient aussi variables au cours du temps : essentiellement la fonction (HAQ) à J0, 6 mois, 2 an et 3 ans et majoritairement l'état mental à 1 an.

La variabilité dans les déterminants de l'utilité SF-6D, EQ-5D et des différences entre les deux mesures d'utilité au cours du temps, les difficultés de modélisation linéaire de l'utilité EQ-5D posent question quant à l'utilisation du mapping pour estimer l'utilité à partir d'instruments de mesure de l'état de santé notamment à partir du HAQ.

Enfin, concernant l'étude de l'utilisation des utilités, c'est à dire les changements d'utilité EQ-5D et SF-6D, nous avons montré qu'ils reflétaient bien l'évolution de l'état de santé des patients souffrant de PR débutante, avec des moyennes de changement et une variabilité toujours plus importantes pour l'EQ-5D. Pour une même stratégie thérapeutique, le gain de QALYs est quasiment doublé avec l'EQ-5D par rapport au SF-6D, ce qui fournira des rapports coût-utilité plus favorables. Cependant la variabilité qui caractérise ce gain de QALYs est elle aussi doublée, ce qui fait que le rapport coût-utilité sera entouré de plus d'imprécision.

Actuellement, il n'y a pas de consensus sur un seuil coût-utilité acceptable définissant un traitement coût-efficace. Aucun pays n'applique de seuil clairement défini mais les seuils "implicites" seraient autour de 50 000 $ par QALY gagné pour les Etats-Unis, 30 000 £ pour l'Angleterre et 80 000 € pour les Pays-Bas (106). Bien sûr la décision de l'allocation des ressources ne repose pas uniquement sur ces rapports coût-utilité : le nombre de patients éligibles à ce nouveau traitement, le budget disponible, les priorités des différentes politiques de santé jouent un rôle déterminant.

Nord et al (107) ont suggéré d'utiliser une étendue de seuil coût-utilité plutôt qu'un seuil unique afin de tenir compte du contexte sans modifier la notion de QALY elle-même : cela permettrait de considérer la sévérité de la pathologie, le manque de meilleur traitement alternatif, ou la situation particulière de la fin de vie.

En conclusion, l'EQ-5D et le SF-6D ne sont pas interchangeables dans la polyarthrite débutante : ils ont des échelles, des distributions, une reproductibilité et une sensibilité au changement différentes. Ils ne donnent pas les mêmes résultats surtout quand les patients ont un mauvais état de santé et plus particulièrement une incapacité fonctionnelle importante.
La distribution bimodale de l'EQ-5D, sa moins bonne reproductibilité, le fait qu'il soit moins sensible à l'amélioration, sa plus grande variabilité dans l'estimation du gain d'utilité et donc des QALYs, les problèmes soulevés par la signification d'un état « pire que la mort » (qui n'est pas celle du patient), favoriseraient plutôt l'utilisation du SF-6D dans les essais sur les nouvelles biothérapies dans la polyarthrite débutante.

# PERSPECTIVES

## I – DANS LA POLYARTHRITE RHUMATOÏDE :

### ❖ Conséquences des différences entre EQ-5D et SF-6D :

Je vais m'intéresser aux conséquences des différences entre les 2 instruments dans 2 utilisations possibles des mesures d'utilités:

- Tout d'abord en calculant et comparant les gains d'utilité obtenus avec les 2 instruments à 2 ans et 3 ans dans l'ensemble de la population et dans différentes sous-populations d'intérêt notamment sous différentes stratégies thérapeutiques dans la cohorte ESPOIR.

  - Etude de valorisation des coûts de la maladie en cours
  - Etude sur le caractère optimal de la stratégie thérapeutique à adapter sur un plus long terme
  - Possibilité de modélisation coût-utilité et d'étude de l'impact des outils

- Puis, en essayant de déterminer si l'utilité EQ-5D et/ou SF-6D est un facteur prédictif de la mortalité, de la consommation de soins ou de l'activité professionnelle.

### ❖ Impact des comorbidités sur l'utilité :

La qualité de vie est modifiée par le phénomène de multimorbidité (82) et il apparait nécessaire d'analyser l'impact de plusieurs comorbidités sur l'utilité. La plupart des études qui ont étudiés l'effet de la mutimorbidité sur la qualité de vie étaient transversales (83,108,109) or l'impact des comorbidités peut varier au cours du temps et il est nécessaire de voir comment il affecte les changements d'utilité.

Mon prochain champ d'investigation sera donc l'impact des comorbidités sur la mesure de l'utilité dans la polyarthrite débutante : la présence ou la survenue d'une comorbidité modifie-t-elle l'impact de la polyarthrite sur les utilités EQ-5D, SF-6D et sur les changements d'utilité? Il est souhaitable que l'instrument d'utilité soit le moins influencé possible par la présence d'une comorbidité tout en restant sensible et discriminant à l'échelon de la polyarthrite.

J'utiliserai les données de la cohorte ESPOIR et d'une cohorte de PR établies (COMEDRA).

Ce travail participera au développement de la prise en charge des comorbidités dans les maladies chroniques avec la recherche des moyens d'évaluer l'impact médico-économique.

- **Impact de l'autoévaluation sur l'utilité :**

L'autoévaluation régulière de son état clinique par le patient est actuellement en plein essor du fait du développement de l'implication des patients dans la gestion de leur pathologie (notion de patient partenaire) et de l'évolution à la baisse de la démographie médicale. Je vais m'intéresser à l'impact de l'autoévaluation sur les mesures d'utilité EQ-5D et SF-6D chez des patients suivis pour PR grâce à l'étude COMEDRA. Cet axe de recherche s'inscrit lui aussi pleinement dans la prise en charge des maladies chroniques.

- **Impact culturel et sociétal sur l'utilité :**

Un autre axe important de ma recherche sera l'étude de l'impact culturel et sociétal sur l'utilité : - tout d'abord en comparant les utilités EQ-5D et SF-6D obtenues avec des pondérations issues de différentes populations de référence (française, anglaise, espagnole) dans la cohorte ESPOIR.

- ensuite en utilisant les données de l'étude RAMP (Rheumatoid Arthritis Mexico-Paris) dont le recrutement vient de se terminer. Il s'agit d'une étude transversale visant à comparer les données socio-démographiques, les caractéristiques de la maladie (activité, fonction…), les niveaux d'utilité EQ-5D et SF-6D, les prises en charge thérapeutiques et les connaissances, croyances de 50 patients mexicains et 50 patients français souffrant de polyarthrite rhumatoïde suivis en ambulatoire. Cela va permettre de voir l'impact culturel, sociétal, des connaissances et des croyances sur les mesures d'utilité EQ-5D et SF-6D.

## I I– AUTRES MALADIES RHUMATOLOGIQUES :

### II.1 – Spondylarthrite ankylosante

❖ **Cohorte DESIR :**

Il s'agit d'une cohorte de 708 patients ayant une spondylarthrite certaine ou probable de moins de 3 ans d'évolution (110). Le score de qualité de vie SF36 a été recueilli permettant l'obtention de l'utilité SF-6D. Je m'intéresserai à l'analyse des propriétés psychométriques, des déterminants et des facteurs prédictifs de l'utilité SF-6D dans la spondylarthrite.

❖ **Protocole ancillaire au PHRC SPACING :**

Il s'agit d'un essai contrôlé randomisé comparant une stratégie d'espacement d'un anti-TNF versus administration habituelle chez les patients souffrants de spondylarthrite en faible activité. L'utilité EQ-5D et SF-6D pourront être déterminées.
Je pourrai ainsi analyser et comparer leurs propriétés psychométriques et leurs déterminants dans la spondylarthrite ankylosante. Je m'intéresserai également à l'influence de ces deux instruments sur le rapport coût-utilité de la comparaison des deux stratégies.

### II.2 – L'arthrose

❖ **Cohorte KHOALA (Knee and Hip OsteoArthritis Long term Assessment):**

Il s'agit d'une cohorte de 878 personnes âgées de 40 à 75 ans souffrant d'arthrose symptomatique de hanche ou de genou (111). Je m'intéresserai à l'analyse et à la comparaison des propriétés psychométriques, des déterminants et des facteurs prédictifs de l'utilité SF-6D et EQ-5D dans l'arthrose symptomatique des membres inférieurs.

Les perspectives de ce travail sont les suivantes :

1- Quelle est l'influence de ces deux instruments sur le rapport coût-utilité de la comparaison de deux stratégies qu'elles soient issues de données de cohorte, avec des stratégies d'intensification chez des patients souffrant de maladies actives (cohorte ESPOIR), ou de données d'essai contrôlé randomisé avec des stratégies de diminution des traitements chez des patients en faible activité ou rémission (PHRC SPACING)?

2- Ces résultats sont-ils pathologie dépendante ou pas ?

# REFERENCES


1. Guillemin F, Saraux A, Guggenbuhl P, Roux CH, Fardellone P, Le Bihan E, et al. Prevalence of rheumatoid arthritis in France: 2001. Ann. Rheum. Dis. 2005 oct;64(10):1427-30.

2. Alamanos Y, Voulgari PV, Drosos AA. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin. Arthritis Rheum. 2006 déc;36(3):182-8.

3. van der Heijde DM. Joint erosions and patients with early rheumatoid arthritis. Br J Rheumatol. 1995;34(Suppl 2):74-8.

4. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham, C. O. 3rd, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010;69(9):1580-8.

5. Albers JM, Kuper HH, van Riel PL, Prevoo ML, van 't Hof MA, van Gestel AM, et al. Socio-economic consequences of rheumatoid arthritis in the first years of the disease. Rheumatology (Oxford). 1999;38(5):423-30.

6. Strand V, Khanna D. The impact of rheumatoid arthritis and treatment on patients' lives. Clin. Exp. Rheumatol. 2010 juin;28(3 Suppl 59):S32-40.

7. Strand V, Sharp V, Koenig AS, Park G, Shi Y, Wang B, et al. Comparison of health-related quality of life in rheumatoid arthritis, psoriatic arthritis and psoriasis and effects of etanercept treatment. Ann. Rheum. Dis. 2012 juill;71(7):1143-50.

8. Dadoun S, Lucier S, De Rosa M, Saraux A, Berenbaum F, Durand-Zaleski I, et al. Rheumatoid Arthritis (RA) Impact on Work and RA-Related Indirect Costs during the First Three Years after RA Diagnosis. An Economic Analysis from the ESPOIR Cohort. ACR 2010. Atlanta (GA); 2010.

9. Hallert E, Björk M, Dahlström O, Skogh T, Thyberg I. Disease activity and disability in women and men with early rheumatoid arthritis (RA): An 8-year followup of a Swedish early RA project. Arthritis Care Res (Hoboken). 2012 août;64(8):1101-7.

10. Minaur NJ, Jacoby RK, Cosh JA, Taylor G, Rasker JJ. Outcome after 40 years with rheumatoid arthritis: a prospective study of function, disease activity, and mortality. J Rheumatol Suppl. 2004;69:3-8.

11. Callahan LF, Pincus T. Mortality in the rheumatic diseases. Arthritis Care Res. 1995;8(4):229-41.

12. Kvalvik AG, Jones MA, Symmons DP. Mortality in a cohort of Norwegian patients with rheumatoid arthritis followed from 1977 to 1992. Scand J Rheumatol. 2000;29(1):29-37.

13. Krause D, Schleusser B, Herborn G, Rau R. Response to methotrexate treatment is associated with reduced mortality in patients with severe rheumatoid arthritis. Arthritis Rheum. 2000;43(1):14-21.

14. Goodson N, Marks J, Lunt M, Symmons D. Cardiovascular admissions and mortality in an inception cohort of patients with rheumatoid arthritis with onset in the 1980s and 1990s. Ann Rheum Dis. 2005;64(11):1595-601. Epub 2005 Apr 20.

15. Goodson NJ, Wiles NJ, Lunt M, Barrett EM, Silman AJ, Symmons DP. Mortality in early inflammatory polyarthritis: cardiovascular mortality is increased in seropositive patients. Arthritis Rheum. 2002;46(8):2010-9.

16. Goodson NJ, Symmons DP, Scott DG, Bunn D, Lunt M, Silman AJ. Baseline levels of C-reactive protein and prediction of death from cardiovascular disease in patients with inflammatory polyarthritis: a ten-year followup study of a primary care-based inception cohort. Arthritis Rheum. 2005;52(8):2293-9.

17. Shoenfeld Y, Gerli R, Doria A, Matsuura E, Cerinic MM, Ronda N, et al. Accelerated atherosclerosis in autoimmune rheumatic diseases. Circulation. 2005;112(21):3337-47.

18. Aviña-Zubieta JA, Choi HK, Sadatsafavi M, Etminan M, Esdaile JM, Lacaille D. Risk of cardiovascular mortality in patients with rheumatoid arthritis: a meta-analysis of observational studies. Arthritis Rheum. 2008 déc 15;59(12):1690-7.

19. Avina-Zubieta JA, Thomas J, Sadatsafavi M, Lehman AJ, Lacaille D. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. Annals of the rheumatic diseases [Internet]. 2012 mars 16 [cité 2012 juill 31]; Available de: http://www.ncbi.nlm.nih.gov/pubmed/22425941

20. Södergren A, Karp K, Boman K, Eriksson C, Lundström E, Smedby T, et al. Atherosclerosis in early rheumatoid arthritis: very early endothelial activation and rapid progression of intima media thickness. Arthritis Res. Ther. 2010;12(4):R158.

21. Choi HK, Hernan MA, Seeger JD, Robins JM, Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. Lancet. 2002;359(9313):1173-7.

22. Dixon WG, Symmons DP. What effects might anti-TNFalpha treatment be expected to have on cardiovascular morbidity and mortality in rheumatoid arthritis? A review of the role of TNFalpha in cardiovascular pathophysiology. Ann Rheum Dis. 2007;66(9):1132-6.

23. Gaujoux-Viala C, Smolen JS, Landewé R, Dougados M, Kvien TK, Mola EM, et al. Current evidence for the management of rheumatoid arthritis with synthetic disease-modifying antirheumatic drugs: a systematic literature review informing the EULAR recommendations for the management of rheumatoid arthritis. Ann. Rheum. Dis. 2010 juin;69(6):1004-9.

24. Huizinga TW, Landewe RB. Early aggressive therapy in rheumatoid arthritis: a « window of opportunity »? Nat Clin Pract Rheumatol. 2005;1(1):2-3.

25. Munro R, Hampson R, McEntegart A, Thomson EA, Madhok R, Capell H. Improved functional outcome in patients with early rheumatoid arthritis treated with

intramuscular gold: results of a five year prospective study. Ann. Rheum. Dis. 1998 févr;57(2):88-93.

26. Lukas C, Combe B, Ravaud P, Sibilia J, Landew R, van der Heijde D. Favorable effect of very early disease-modifying antirheumatic drug treatment on radiographic progression in early inflammatory arthritis: Data from the Etude et Suivi des polyarthrites indifferenciees recentes (study and followup of early undifferentiated polyarthritis). Arthritis Rheum. 2011;63(7):1804-11.

27. van der Heide A, Jacobs JW, Bijlsma JW, Heurkens AH, van Booma-Frankfort C, van der Veen MJ, et al. The effectiveness of early treatment with « second-line » antirheumatic drugs. A randomized, controlled trial. Ann Intern Med. 1996;124(8):699-707.

28. Anderson JJ, Wells G, Verhoeven AC, Felson DT. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. Arthritis Rheum. 2000;43(1):22-9.

29. Nell VP, Machold KP, Eberl G, Stamm TA, Uffmann M, Smolen JS. Benefit of very early referral and very early therapy with disease-modifying anti-rheumatic drugs in patients with early rheumatoid arthritis. Rheumatology (Oxford). 2004;43(7):906-14.

30. van Aken J, Lard LR, le Cessie S, Hazes JMW, Breedveld FC, Huizinga TWJ. Radiological outcome after four years of early versus delayed treatment strategy in patients with recent onset rheumatoid arthritis. Ann. Rheum. Dis. 2004 mars;63(3):274-9.

31. Emery P, Marzo H, Proudman S. Management of patients with newly diagnosed rheumatoid arthritis. Rheumatology (Oxford). 1999;38(Suppl 2):27-31.

32. Emery P. Evidence supporting the benefit of early intervention in rheumatoid arthritis. J Rheumatol (Suppl). 2002;66:3-8.

33. Lard LR, Visser H, Speyer I, vander Horst-Bruinsma IE, Zwinderman AH, Breedveld FC, et al. Early versus delayed treatment in patients with recent-onset rheumatoid arthritis: comparison of two cohorts who received different treatment strategies. Am J Med. 2001;111(6):446-51.

34. Franke LC, Ament AJ, van de Laar MA, Boonen A, Severens JL. Cost-of-illness of rheumatoid arthritis and ankylosing spondylitis. Clin Exp Rheumatol. 2009;27(4 Suppl 55):S118-23.

35. Mau W, Listing J, Huscher D, Zeidler H, Zink A. Employment across chronic inflammatory rheumatic diseases and comparison with the general population. J Rheumatol. 2005;32(4):721-8.

36. Bjork M, Thyberg I, Rikner K, Balogh I, Gerdle B. Sick leave before and after diagnosis of rheumatoid arthritis–a report from the Swedish TIRA project. J Rheumatol. 2009;36(6):1170-9.

37. Xie F. The need for standardization: a literature review of indirect costs of rheumatoid arthritis and osteoarthritis. Arthritis Rheum. 2008;59(7):1027-33.

38. Fautrel B, Guillemin F. Cost of illness studies in rheumatic diseases. Curr Opin Rheumatol. 2002;14(2):121-6.

39. Fautrel B, Penrod JR, Clarke AE. The economics of musculoskeletal diseases. In: Isenberg DA, Maddison PJ, Woo P, Glass D, Breedveld FC, éditeurs. Oxford Textbook of Rheumatology. 3rd éd. Oxford: Oxford University Presse; 2004. p. 285-296.

40. Drummond M, Maetzel A, Gabriel S, March L. Towards a reference case for use in future economic evaluations of interventions in osteoarthritis. J Rheumatol Suppl. 2003;68:26-30.

41. Merkesdal S, Ruof J, Huelsemann JL, Schoeffski O, Maetzel A, Mau W, et al. Development of a matrix of cost domains in economic evaluation of rheumatoid arthritis. J Rheumatol. 2001;28(3):657-61.

42. Gabriel S, Drummond M, Suarez-Almazor ME, Ruff B, Guillemin F, Bombadier C, et al. OMERACT 5 Economics Working Group: summary, recommendations, and research agenda. J Rheumatol. 2001;28(3):670-3.

43. Gabriel S, Drummond M, Maetzel A, Boers M, Coyle D, Welch V, et al. OMERACT 6 Economics Working Group report: a proposal for a reference case for economic evaluation in rheumatoid arthritis. J Rheumatol. 2003;30(4):886-90.

44. Guillemin F, Durieux S, Daures JP, Lafuma A, Saraux A, Sibilia J, et al. Costs of rheumatoid arthritis in France: a multicenter study of 1109 patients managed by hospital-based rheumatologists. J Rheumatol. 2004;31(7):1297-304.

45. Kobelt G, Woronoff AS, Richard B, Peeters P, Sany J. Disease status, costs and quality of life of patients with rheumatoid arthritis in France: the ECO-PR Study. Joint Bone Spine. 2008;75(4):408-15.

46. Fautrel B, Verstappen SMM, Boonen A. Economic consequences and potential benefits. Best Pract Res Clin Rheumatol. 2011 août;25(4):607-24.

47. Momohara S, Inoue E, Ikari K, Kawamura K, Tsukahara S, Iwamoto T, et al. Decrease in orthopaedic operations, including total joint replacements, in patients with rheumatoid arthritis between 2001 and 2007: data from Japanese outpatients in a single institute-based large observational cohort (IORRA). Ann Rheum Dis. 2010;69(1):312-3.

48. Sokka T, Kautiainen H, Pincus T, Verstappen SMM, Aggarwal A, Alten R, et al. Work disability remains a major problem in rheumatoid arthritis in the 2000s: data from 32 countries in the QUEST-RA study. Arthritis Res. Ther. 2010;12(2):R42.

49. Hallert E, Husberg M, Bernfort L. The incidence of permanent work disability in patients with rheumatoid arthritis in Sweden 1990-2010: before and after introduction of biologic agents. Rheumatology (Oxford). 2012 févr;51(2):338-46.

50. Olofsson T, Englund M, Saxne T, Jöud A, Jacobsson LTH, Geborek P, et al. Decrease in sick leave among patients with rheumatoid arthritis in the first 12 months after start of treatment with tumour necrosis factor antagonists: a population-based controlled cohort study. Ann. Rheum. Dis. 2010 déc;69(12):2131-6.

51. Kernick DP. Introduction to health economics for the medical practitioner. Postgrad Med J. 2003;79(929):147-50.

52. Drummond MF, Sculpher MJ, Torrance GW. Methods for economic evaluation of health care programmes. Oxford: Oxford Medical Publications; 2005.

53. O'Brien B. Principles of economic evaluation for health care programs. J Rheumatol. 1995;22(7):1399-402.

54. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. Pharmacoeconomics. 1998;13(4):397-409.

55. Rice DP. Cost-of-illness studies: fact or fiction? Lancet. 1994;344(8936):1519-20.

56. Coyle D, Welch V, Shea B, Gabriel S, Drummond M, Tugwell P. Issues of consensus and debate for economic evaluation in rheumatology. J Rheumatol. 2001;28(3):642-7.

57. Gabriel SE, Tugwell P, Drummond M. Progress towards an OMERACT-ILAR guideline for economic evaluations in rheumatology. Ann Rheum Dis. 2002;61(4):370-3.

58. Bridges JF. Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. Appl Health Econ Health Policy. 2003;2(4):213-24.

59. Khanna D, Tsevat J. Health-related quality of life--an introduction. Am J Manag Care. 2007 déc;13 Suppl 9:S218-223.

60. Neumann J von, Morgenstern O. Theory of Games and Economic Behavior (Commemorative Edition). Princeton University Press; 2007.

61. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res. 1972;7(2):118-33.

62. Bell D, Farquhar P. Perspectives on utility theory. Operations Research. 1986;179-83.

63. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N. Engl. J. Med. 1977 mars 31;296(13):716-21.

64. Williams A. Economics of coronary artery bypass grafting. Br Med J (Clin Res Ed). 1985 août 3;291(6491):326-9.

65. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986 mars;5(1):1-30.

66. Bakker C, van der Linden S. Health related utility measurement: an introduction. J. Rheumatol. 1995 juin;22(6):1197-9.

67. Puhan MA, Schünemann HJ, Wong E, Griffith L, Guyatt GH. The standard gamble showed better construct validity than the time trade-off. J Clin Epidemiol. 2007 oct;60(10):1029-33.

68. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 1987.

69. Froberg DG, Kane RL. Methodology for measuring health-state preferences--II: Scaling methods. J Clin Epidemiol. 1989;42(5):459-71.

70. Dolan P, Gudex C. Time preference, duration and health state valuations. Health Econ. 1995 août;4(4):289-99.

71. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990 déc;16(3):199-208.

72. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271-92.

73. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes. 2003;1:54.

74. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 nov;35(11):1095-108.

75. Chevalier J, de Pouvourville G. Valuing EQ-5D using Time Trade-Off in France. Eur J Health Econ. 2011 Sep 21. [Epub ahead of print]

76. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 juin;30(6):473-83.

77. Leplège A, Ecosse E, Verdier A, Perneger TV. The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. J Clin Epidemiol. 1998 nov;51(11):1013-23.

78. Carr-Hill RA. Assumptions of the QALY procedure. Soc Sci Med. 1989;29(3):469-77.

79. Moatti JP, Auquier P, Le Coroller AG, Macquart-Moulin G. [QALYS or not QALYS: that is the question?]. Rev Epidemiol Sante Publique. 1995 déc;43(6):573-83.

80. Robine JM, Mathers CD, Bucquet D. Distinguishing health expectancies and health-adjusted life expectancies from quality-adjusted life years. Am J Public Health. 1993;83(6):797-8.

81. Guillemin F. The value of utility: assumptions underlying preferences and quality adjusted life years. J Rheumatol. 1999;26(9):1861-3.

82. van den Akker M, Buntinx F, Metsemakers JF, Roos S, Knottnerus JA. Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. J Clin Epidemiol. 1998 mai;51(5):367-75.

83. Fortin M, Lapointe L, Hudon C, Vanasse A, Ntetu AL, Maltais D. Multimorbidity and quality of life in primary care: a systematic review. Health Qual Life Outcomes. 2004;2:51.

84. Schwartz CE, Sprangers MAG. Reflections on genes and sustainable change: toward a trait and state conceptualization of response shift. J Clin Epidemiol. 2009 nov;62(11):1118-23.

85. La Puma J, Lawlor EF. Quality-adjusted life-years. Ethical implications for physicians and policymakers. JAMA. 1990 juin 6;263(21):2917-21.

86. Ried W. QALYs versus HYEs--what's right and what's wrong. A review of the controversy. J Health Econ. 1998 oct;17(5):607-25.

87. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. Med Decis Making. 1989 juin;9(2):142-9.

88. Pen CL. Théorie de l'utilité et mesure des états de santé, le débat QALYs-HYEs. Économie & prévision. 129(3):37-54.

89. Xie F, Li S-C, Luo N, Lo N-N, Yeo S-J, Yang K-Y, et al. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. Arthritis Rheum. 2007 août 15;57(6):1043-9.

90. Boonen A, van der Heijde D, Landewé R, van Tubergen A, Mielants H, Dougados M, et al. How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? Ann. Rheum. Dis. 2007 juin;66(6):771-7.

91. Lillegraven S, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. Ann. Rheum. Dis. 2010 oct;69(10):1762-7.

92. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J. Rheumatol. 1998 févr;25(2):198-9.

93. Coste J, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. Stat Med. 1995 févr 28;14(4):331-45.

94. Guillemin F, Billot L, Boini S, Gerard N, Ødegaard S, Kvien TK. Reproducibility and sensitivity to change of 5 methods for scoring hand radiographic damage in patients with rheumatoid arthritis. J. Rheumatol. 2005 mai;32(5):778-86.

95. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 juill;63(7):737-45.

96. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology (Oxford). 2002 déc;41(12):1380-7.

97. Linde L, Sørensen J, Ostergaard M, Hørslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, 15D, EQ-5D [corrected] RAQoL, and HAQ in patients with rheumatoid arthritis. J. Rheumatol. 2008 août;35(8):1528-37.

98. Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Farragher TM, Verstappen SMM, et al. Why do patients with inflammatory arthritis often score states « worse than death » on the EQ-5D? An Investigation of the EQ-5D classification system. Value Health. 2009 sept;12(6):1026-34.

99. Adams R, Walsh C, Veale D, Bresnihan B, FitzGerald O, Barry M. Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics. 2010;28(6):477-87.

100. Macran S, Kind P. « Death » and the valuation of health-related quality of life. Med Care. 2001 mars;39(3):217-27.

101. De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ. 2000 mars;9(2):109-26.

102. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med. 1999 juin;48(11):1507-15.

103. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med. 1999 juin;48(11):1531-48.

104. Sprangers MA, Van Dam FS, Broersen J, Lodder L, Wever L, Visser MR, et al. Revealing response shift in longitudinal research on fatigue--the use of the thentest approach. Acta Oncol. 1999;38(6):709-18.

105. Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities implications for policy. Med Care. 1999 mai;37(5):479-89.

106. Eichler H-G, Kong SX, Gerth WC, Mavros P, Jönsson B. Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? Value Health. 2004 oct;7(5):518-28.

107. Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health. 2009 mars;12 Suppl 1:S10-15.

108. Loza E, Jover JA, Rodriguez L, Carmona L. Multimorbidity: prevalence, effect on quality of life and daily functioning, and variation of this effect when one condition is a rheumatic disease. Semin. Arthritis Rheum. 2009 févr;38(4):312-9.

109. Taylor AW, Price K, Gill TK, Adams R, Pilkington R, Carrangis N, et al. Multimorbidity - not just an older person’s issue. Results from an Australian biomedical study. BMC Public Health. 2010;10:718.

110. Dougados M, d’ Agostino M-A, Benessiano J, Berenbaum F, Breban M, Claudepierre P, et al. The DESIR cohort: a 10-year follow-up of early inflammatory back pain in France: study design and baseline characteristics of the 708 recruited patients. Joint Bone Spine. 2011 déc;78(6):598-603.

111. Guillemin F, Rat A-C, Roux CH, Fautrel B, Mazieres B, Chevalier X, et al. The KHOALA cohort of knee and hip osteoarthritis in France. Joint Bone Spine. 2012 [Epub ahead of print]

# ANNEXES

## Annexe A :

## EQ-5D

Veuillez indiquer, pour chacune des rubriques suivantes, l'affirmation qui décrit le mieux votre état de santé aujourd'hui, en cochant la case appropriée.

**Mobilité**

Je n'ai aucun problème pour me déplacer à pied

J'ai des problèmes pour me déplacer à pied

Je suis obligé(e) de rester alité(e)

**Autonomie de la personne**

Je n'ai aucun problème pour prendre soin de moi

J'ai des problèmes pour me laver ou m'habiller tout(e) seul(e)

Je suis incapable de me laver ou de m'habiller tout(e) seul(e)

**Activités courantes** (*exemples : travail, études, travaux domestiques, activités familiales ou loisirs)*

Je n'ai aucun problème pour accomplir mes activités courantes

J'ai des problèmes pour accomplir mes activités courantes

Je suis incapable d'accomplir mes activités courantes

**Douleurs/gêne**

Je n'ai ni douleurs, ni gêne

J'ai des douleurs ou une gêne modérée(s)

J'ai des douleurs ou une gêne extrême(s)

**Anxiété/dépression**

Je ne suis ni anxieux(se), ni déprimé(e)

Je suis modérément anxieux(se) ou déprimé(e)

Je suis extrêmement anxieux(se) ou déprimé(e)

**Annexe B** : le SF-36

Les questions qui suivent portent sur votre santé, telle que vous la ressentez. Ces informations nous permettront de mieux savoir comment vous vous sentez dans votre vie de tous les jours.

***Veuillez répondre à toutes les questions en cochant la case correspondant à la réponse choisie, comme il est indiqué. Si vous ne savez pas très bien comment répondre, choisissez la réponse la plus proche de votre situation.***

1. **Dans l'ensemble, pensez-vous que votre santé est :**

cochez la case qui correspond à votre choix

- Excellente..... ❑
- Très bonne..... ❑
- Bonne ..... ❑
- Médiocre ..... ❑
- Mauvaise ..... ❑

2. **Par rapport à il y a 6 mois, comment trouvez-vous votre état de santé en ce moment ?**

cochez la case qui correspond à votre choix

- Bien meilleur qu'il y a 6 mois ..... ❑
- Plutôt meilleur ..... ❑
- A peu près pareil..... ❑
- Plutôt moins bon..... ❑
- Beaucoup moins bon ..... ❑

3. ***Voici une liste d'activités que vous pouvez avoir à faire dans votre vie de tous les jours. Pour chacune d'entre elles indiquez si vous êtes limité(e) en raison de votre état de santé actuel.***

cochez la case qui correspond à votre choix

| Liste d'activités | **oui, beaucoup limité(e)** | **oui, un peu limité(e)** | **non, pas du tout limité(e)** |
|---|---|---|---|
| a. **Efforts physiques importants** tels que courir, soulever un objet lourd, faire du sport | ❑ | ❑ | ❑ |
| b. **Efforts physiques modérés** tels que déplacer une table, passer l'aspirateur, jouer aux boules | ❑ | ❑ | ❑ |
| c. Soulever et porter les courses | ❑ | ❑ | ❑ |
| d. Monter **plusieurs étages** par l'escalier | ❑ | ❑ | ❑ |
| e. Monter **un étage** par l'escalier | ❑ | ❑ | ❑ |
| f. Se pencher en avant, se mettre à genoux, s'accroupir | ❑ | ❑ | ❑ |
| g. Marcher **plus d'un km** à pied | ❑ | ❑ | ❑ |
| h. Marcher **plusieurs centaines de mètres** | ❑ | ❑ | ❑ |
| i. Marcher **une centaine de mètres** | ❑ | ❑ | ❑ |
| j. Prendre un bain, une douche ou s'habiller | ❑ | ❑ | ❑ |

4. ***Au cours de <u>ces 8 derniers jours</u>, et en raison de votre état physique,***

cochez la case qui correspond à votre choix

| | **OUI** | **NON** |
|---|---|---|
| a. Avez-vous réduit **le temps passé** à votre travail ou à vos activités habituelles ? | ❑ | ❑ |
| b. Avez-vous **accompli moins** de choses que vous auriez souhaitées ? | ❑ | ❑ |
| c. Avez-vous dû arrêter de faire **certaines choses ?** | ❑ | ❑ |
| d. Avez-vous eu des **difficultés** à faire votre travail ou toute autre activité (par exemple, cela vous a demandé un effort supplémentaire) ? | ❑ | ❑ |

5. ***Au cours de ces <u>8 derniers jours</u>, et en raison de votre état émotionnel (comme vous sentir triste, nerveux(se) ou déprimé(e)),***

cochez la case qui correspond à votre choix

| | **OUI** | **NON** |
|---|---|---|
| a. Avez-vous réduit **le temps passé** à votre travail ou à vos activités habituelles ? | ❑ | ❑ |
| b. Avez-vous **accompli moins** de choses que vous auriez souhaitées ? | ❑ | ❑ |
| c. Avez-vous eu des **difficultés** à faire ce que vous aviez à faire **avec autant de soin et d'attention** que d'habitude ? | ❑ | ❑ |

6. ***Au cours de ces <u>8 derniers jours</u> dans quelle mesure votre état de santé, physique ou émotionnel, vous a-t-il gêné(e) dans votre vie sociale et vos relations avec les autres, votre famille, vos amis, vos connaissances ?***

cochez la case qui correspond à votre choix

- Pas du tout ..... ❑
- Un petit peu ..... ❑
- Moyennement ..... ❑
- Beaucoup ..... ❑
- Enormément..... ❑

7. ***Au cours de ces <u>8 derniers jours</u>, quelle a été l'intensité de vos <u>douleurs physiques</u> ?***

cochez la case qui correspond à votre choix

- Nulle ..... ❑
- Très faible ..... ❑
- Faible ..... ❑
- Moyenne ..... ❑
- Grande ..... ❑
- Très grande ..... ❑

8. ***Au cours de <u>ces 8 derniers jours</u>, dans quelle mesure vos douleurs physiques vous ont-elles limité(e) dans votre travail ou vos activités domestiques ?***

cochez la case qui correspond à votre choix

- Pas du tout ..... ❑
- Un petit peu ..... ❑
- Moyennement ..... ❑
- Beaucoup ..... ❑
- Enormément ..... ❑

9. ***Les questions qui suivent portent sur comment vous vous êtes senti(e) au cours de ces <u>8 derniers jours</u>. Pour chaque question, veuillez indiquer la réponse qui vous semble la plus appropriée. Au cours de ces 8 derniers jours, y a-t-il eu des moments où :***

cochez la case qui correspond à votre choix

| | **en permanence** | **très souvent** | **souvent** | **quelquefois** | **rarement** | **jamais** |
|---|---|---|---|---|---|---|
| a. Vous vous êtes senti(e) dynamique ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| b. Vous vous êtes senti(e) très nerveux (se) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| c. Vous êtes vous senti(e) si découragé(e) que rien ne pouvait vous remonter le moral ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| d. Vous êtes vous senti(e) calme et détendu(e) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| e. Vous êtes vous senti(e) débordant(e) d'énergie ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| f. Vous êtes vous senti(e) triste et abattu(e) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| g. Vous êtes vous senti(e) épuisé(e) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| h. Vous êtes vous senti(e) heureux (se) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |
| i. Vous êtes vous senti(e) fatigué(e) ? | ❑ | ❑ | ❑ | ❑ | ❑ | ❑ |

10. ***Au cours de ces <u>8 derniers jours</u> y a-t-il eu des moments où <u>votre état de santé, physique ou émotionnel</u>, vous a gêné(e) dans votre vie sociale et vos relations avec les autres, votre famille, vos amis, vos connaissances ?***

cochez la case qui correspond à votre choix

- En permanence ..... ❑
- Une bonne partie du temps..... ❑
- De temps en temps ..... ❑
- Rarement ..... ❑
- Jamais ..... ❑

11. ***Indiquez, pour chacune des phrases suivantes, dans quelle mesure elles sont vraies ou fausses dans votre cas :***

cochez la case qui correspond à votre choix

| | **totalement vraie** | **plutôt vraie** | **je ne sais pas** | **plutôt fausse** | **totalement fausse** |
|---|---|---|---|---|---|
| a. Je tombe malade plus facilement que les autres | ❑ | ❑ | ❑ | ❑ | ❑ |
| b. Je me porte aussi bien que n'importe qui | ❑ | ❑ | ❑ | ❑ | ❑ |
| c. Je m'attends à ce que ma santé se dégrade | ❑ | ❑ | ❑ | ❑ | ❑ |
| d. Je suis en excellente santé | ❑ | ❑ | ❑ | ❑ | ❑ |

# Who are the patients with early arthritis with "worse than death" scores on the EQ-5D? Results from the ESPOIR cohort

Cécile Gaujoux-Viala[12], Bruno Fautrel[2], Francis Guillemin[13], René-Marc Flipo[4], Pierre Bourgeois[2], Anne-Christine Rat[13]

**Affiliations:**

[1] Lorraine University, Paris Descartes University, EA 4360 Apemac, Nancy, France

[2] Paris 6 – Pierre et Marie Curie University; AP-HP, Department of Rheumatology, Pitié-Salpêtrière Hospital, Paris, France

[3] INSERM, CIC-EC CIE6, Nancy, France

[4] Department of Rheumatology, Lille University 2, France

**Institution to which the work should be attributed**: Lorraine University, Paris Descartes University, EA 4360 Apemac; Paris 6 – Pierre et Marie Curie University; Department of Rheumatology, Pitié-Salpêtrière Hospital, Paris, France

**Authors:**

Cécile Gaujoux-Viala, MD, Anne-Christine Rat, MD, PhD, Francis Guillemin, MD, PhD, René-Marc Flipo, MD, PhD, Patrice Fardellone, MD, PhD, Pierre Bourgeois MD, Bruno Fautrel, MD, PhD

**Corresponding author**: Dr Cécile Gaujoux-Viala, corresponding author
Université Paris 6 – Pierre & Marie Curie
AP-HP, Groupe hospitalier Pitié-Salpêtrière,
Service de Rhumatologie,
83 boulevard de l'Hôpital 75651 Paris cedex 13, France

Phone: +33 1.421.77821 – Fax : +33.1.421.77955

Email: cecilegaujouxviala@yahoo.fr

**Short running footline**: Health state worse than death in early arthritis

**Words count: 2728**

**Abstract**. 246 words

Objectives: The EQ-5D is a 5-dimension multi-attribute questionnaire widely used to calculate quality-adjusted life-year scores. We aimed to describe patients with early arthritis (EA) and EQ-5D<0, a state "worse than death", at baseline and over 2 years and determine aspects of EA associated with a negative score.

Methods: EQ-5D scores for 813 patients were longitudinally assessed over 24 months. Characteristics and health status of patients with EQ-5D<0 were analyzed. Multivariate logistic regression was used to determine aspects of EA associated with a negative score.

Results: At baseline, 90 (11%) patients had a negative EQ-5D score (median EQ-5D -0.052; range -0.530 to -0.011). Almost all patients had extreme pain/discomfort and at least moderate problems in anxiety/depression domain. These patients had lower educational level, greater disease activity, higher Health Assessment Questionnaire (HAQ) score (greater functional disability), greater pain and fatigue, lower SF-36 mental component scores and higher C-reactive protein level than patients with EQ-5D≥0 (all p<0.001). In the logistic regression model, worse HAQ and SF-36 mental component scores were associated with EQ-5D<0. Results were similar at 6, 12 and 24 months, with 25, 23 and 24 patients, respectively, in a state "worse than death".

Conclusion: Pain/discomfort, often severe, is a key domain of the EQ-5D for patients with EA in a health state "worse than death", and poor HAQ and SF-36 mental component scores were always independently associated with a state "worse than death". EA patients may have specific needs necessitating better management of pain and psychological state.

**Key words**: utility, EQ-5D, early arthritis, pain

## Introduction

Given the scarcity of health care resources, public and private agencies are interested in the effectiveness and cost-effectiveness of health care interventions [1]. The preferred approach to measuring benefits in cost-effectiveness analysis is to measure health status by a single unit known as utilities. Utility values are used to calculate quality-adjusted life years (QALYs), which are outcome measures that combine quantity and quality of life [2]. QALYs are the number of years of life in which a person is in a perfect health state [3]. They are considered equivalent to years actually lived in a given health state multiplied by the utility value of this state. QALYs are used to assess the benefit of the health interventions in economic evaluations, such as cost-utility analysis. They are particularly informative for health policy decisions because they allow for direct comparison of the efficiency of health care resource expenditure across a wide range of conditions and treatments.

The utility concept was developed by health economists. Assessment of utility assigns a numeric value from 0 to 1 for health states, 0 indicating death and 1 a state of perfect health. The values reflect the preference for a health state in a situation of choice that includes uncertainty or sacrifice (e.g., life years). Methods such as standard gamble and time trade-off may be used to measure health states directly, but they are less suitable for clinical research and less widely used for feasibility reasons. Therefore, indirect utility assessment techniques such as the EQ-5D [4] and SF-6D [5] have been developed. Indirect health utility assessments involve population-assigned weights to calculate utility scores for particular health states from multidomain health-status questionnaires completed by patients [6]. These preference-based measures of health have become an important set of instruments for estimating health state values to calculate QALYs.

The EQ-5D [4] is an indirect preference-based health-related quality of life (HRQoL) instrument increasingly being used for economic evaluation of clinical interventions and health programs. The EQ-5D consists of 5 items or dimensions (mobility, self-care, usual activity, pain/discomfort, and

anxiety/depression). Each dimension has 3 answer categories: 1, no problem, 2, some problems, or 3, severe problems. The combination of the 5 dimensions with the 3 answering categories produces 243 possible health states described as vectors ranging from 11111 (no problems in any domain) to 33333 (severe problems in all 5 domains) (e.g., 11231 corresponds to no problems walking around, no problems with self-care, some problems with performing usual activities, extreme pain or discomfort and not anxious or depressed). The EQ-5D can be reported in terms of the 5-digit profile representing the score on each dimension or in terms of a preference-based, single index number. The latter is calculated by applying algorithms that link the 5-digit health state description to average values for members of the general population by the time trade-off method. The EQ-5D allows for negative utility values, which theoretically correspond to health states "worse than death," not available with other indirect utility instruments (like SF-6D…) [4].

We aimed to describe patients with EQ-5D < 0, a health state "worse than death", in a multi-site cohort of patients with EA, at baseline and over 2 years and to determine the aspects of EA associated with a state "worse than death".

## Patients and Methods.

### *Patients*

Between December 2002 and March 2005, 813 patients with EA from 14 French regional centers were included in the ESPOIR cohort [7]. Inclusion criteria were age 18 to 70 years, more than 2 swollen joints for >6 weeks and <6 months, suspected or confirmed diagnosis of RA, and not taking disease-modifying antirheumatic drugs or steroids (except if <2 weeks before inclusion). Patients were followed every 6 months during the first 2 years. At baseline and at each visit, we collected data on clinical and biological variables, including the Disease Activity Score for 28 joints (DAS28) [8]. Furthermore, at each visit, patients completed self-administered patient-reported outcome questionnaires, including a functional ability questionnaire, the Health Assessment Questionnaire (HAQ) [9], and the HRQoL questionnaires EQ-5D and SF-36 [10]. The SF-36 has 8 dimensions with scores ranging from 0 (the worst HRQoL) to 100 (best HRQoL) summarized in 2 components assessing the physical dimension (physical component summary score: PCS) and mental dimension (MCS), with scores transformed to have a mean of 50 and SD 10 [10]. Low PCS and MCS indicate low physical and mental functioning, respectively. The protocol of the ESPOIR cohort study was approved by the ethics committee of Montpellier, France (approval number 020307). All patients gave their signed informed consent before inclusion.

### *Utility measurement*

The utility is calculated by applying algorithms that link the 5-digit health state description to average values for members of the general population by the time trade-off method: 0 indicates death, 1 a state of perfect health, negative (<0) utility indicates a state “worse than death”. EQ-5D reference indices were the French EQ-5D values derived from a population survey in France and involving 10-year time trade-off values [11]. Utility scores generated by the EQ-5D range from -0.53 to 1.

*Statistical analysis*

*Utility scores and profiles of patients with a health state worse than death*

We computed descriptive statistics (median, minimum, maximum and distribution) for patients with EQ-5D < 0 at baseline and the 5-digit ED-5D profile scores for the 5 dimensions.

We investigated whether the mobility, pain or discomfort, and anxiety or depression domains of the EQ-5D are valid measures of functional state, pain and psychological state in EA patients with EQ-5D < 0. We calculated the median HAQ score and PCS of the SF-36 for each of the 3 levels of the mobility domain; median pain at rest and during activities for each of the 3 levels of the pain/ discomfort domain; and median MCS of the SF-36 for each of the 3 levels of the anxiety/depression domain. Kruskal-Wallis test was used to evaluate differences among findings.

*Determinants of health state worse than death*

Multivariate logistic regression was used to determine specific aspects of EA independently associated with EQ-5D < 0 at baseline.

We compared the characteristics and health status of patients with EQ-5D < 0 and EQ-5D ≥ 0 at baseline: qualitative variables were compared by chi-square test or Fisher's exact test as appropriate and quantitative variables with one-way ANOVA or Mann–Whitney U-test as appropriate.

A logistic regression model was then constructed with the dependent variable EQ-5D < 0 (yes/no) at baseline. The model entered candidate variables with a p value <0.1 on bivariate analysis. If several outcomes had similar underlying constructs, only one was selected to avoid collinearity. Several selection procedures (backward elimination, stepwise selection) were tested to check the robustness of the models. All analyses were repeated at 6, 12, and 24 months to check the stability of the model.

Analyses involved use of SAS v9.1 (SAS Inst., Cary, NC, USA). A $p<0.05$ was considered statistically significant.

## Results

*Characteristics of the population*

Table 1 shows the demographic and clinical characteristics of the 813 patients in the ESPOIR cohort at inclusion. In total, 641 (78.8%) patients fulfilled the 2010 criteria of the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) for rheumatoid arthritis.

*Utility scores*

At baseline, 90 (11%) patients showed EQ-5D < 0, a health state "worse than death". Median EQ-5D value was -0.052 (range -0.530 to -0.011). The number of patients at each level of the range gradually increased, with most patients having a score between -0.1 and 0 (Figure 1). Among 813 patients, there were 5 missing value for EQ-5D (14 for the DAS28, 2 for the patient global assessment…).

*Profile of patients with health state "worse than death"*

Patients exhibited 14 health profiles. Almost half of the patients with EQ-5D < 0 (46.7%) had the health profile 22232: some or moderate problems with mobility, self-care, usual activities, extreme pain/discomfort, and some or moderate anxiety/depression. The utility score for this state is -0.052. Almost all patients with EQ-5D < 0 had extreme pain/discomfort and at least moderate problems in usual activity (Figure 2). Fifty-three (59%) of these patients reported moderate and 36 (40%) extreme anxiety/ depression. Most patients reported moderate problems in mobility (92%) and in

self-care (91%). The same distribution of profiles was found at 6, 12 and 24 months, with 25 (3.3%), 23 (3.2%) and 24 (3.5%) patients, respectively, with EQ-5D < 0 (Figure 3). Of the 25 patients with a negative EQ-5D score at 6 months, 12 had a negative score at baseline. Of the 23 patients with a negative EQ-5D score at 12 months, 8 had a negative score at 6 months and only 5 patients had a negative score at all times. Of the 24 patients with a negative EQ-5D score at 2 years, 7 had a negative score at 1 year and 12 at baseline. Only 3 patients had a negative score at all times.

The SF-36 MCS significantly decreased with increasing anxiety/depression score (p=0.005). The SF-36 PCS did not differ among mobility levels (Table 2). The HAQ score increased but not significantly with increasing mobility level (the number of patients was low in some levels). Pain at rest increased with increasing pain/ discomfort score but not significantly, and pain during daily activities did not differ by pain/discomfort level. Thus, the anxiety /depression, mobility and pain/discomfort domains of the EQ-5D may have similar underlying constructs as the SF-36 MCS and, to a lesser extent, the HAQ score and pain at rest for patients with EQ-5D < 0.

*Determinants of health state "worse than death"*

Patients with EQ-5D < 0 had lower educational level, greater disease activity (DAS28), higher HAQ score (greater functional disability), lower SF-36 PCS and MCS, more pain at rest and fatigue, and higher C-reactive protein level than patients with EQ-5D ≥ 0 (all p<0.001) (Table 1). We included level of education, DAS28, HAQ score, SF-36 MCS, pain at rest, fatigue, and CRP level in the multivariate logistic regression model. Increased HAQ score and low SF-36 MCS (poor mental health) were independently associated with negative EQ-5D score at baseline (Table 3). Results were similar at 6, 12 and 24 months and with different variable selection procedures.

## Discussion

This study is the first to examine in a real life setting details of the health states of EA patients with state “worse than death” according to the EQ-5D utility measure and evolution over time. The disease’s impact on pain, functional ability and mental component reflected in a state “worse than death” have consequences for the management of EA in these patients.

This health state was not rare; indeed, 11% of the ESPOIR cohort was in a state “worse than death” at the baseline measurement. This proportion was decreased at 6 months but remained stable over time, at 3.3%. Pain/discomfort was the key EQ-5D domain associated with a state worse than death and was extreme for most patients with EQ-5D < 0, especially at baseline. On multivariate analysis, the HAQ score and SF-36 MCS were independently and consistently associated with a state “worse than death” over 2 years.

One of the strengths of this study is that we investigated data for a large group of patients with EA. The observational ESPOIR cohort study aimed to include patients with EA at all disease levels, age and sex, and here we found a significant proportion of EA patients in a real life setting in a health state “worse than death” by the EQ-5D. The study also included repeated measurement over 2 years with a large number of patients.

Nevertheless it should be noted that the worse than death utility prevalence in a general population matched for age and gender is unknown. The results must be interpreted in light of how the EQ-5D values are generated. The category “worse than death” is a researcher-generated artifact reflecting both the construction of the EQ-5D instrument and the perception of these health states by the general population. Macran et al. identified a small group of subjects from a representative sample in the United Kingdom who, when given the option, choose to place relatively high value on the state "death" as compared with other health states [12]. We do not know whether patients would agree to this definition of their health status, because the patient and population perception of

burden of a disease can be discordant [13]. Investigating further the perception of patients and non-patients of health states by using qualitative research methods would be of interest.

Our results were similar to those found for a more heterogeneous sample of patients with rheumatoid arthritis in a randomised control trial [14]. This previous study investigated baseline data from 2 UK trials: one of patients with inflammatory arthritis of < 11 weeks' symptom duration and the other of patients with established (>5 years) symptomatic RA. However, the 2 groups were not differentiated in the analysis. Of 700 patients included, 9% were in a state "worse than death" and exhibited 9 of the possible 84 health profiles, with more than half (53%) exhibiting the same health profile 22232 we found. All patients presented severe pain/discomfort and moderate problems in 3 or more other domains. Adams et al. [15] found that for 345 patients with established RA, 17% had scores defined as "worse than death" at baseline and 7% at follow-up. Fewer patients with psoriatic arthritis showed this utility (12% at baseline and 2% at follow-up). Extreme pain scores were strongly associated with a state "worse than death".

This decrease of the proportion of patients in state "worse than death" could reflect the beneficial effect of treatment but also an adaptation over time to a new disease state. To note, data were not available to determine if some of the patients with a state "worse than death" had depression or fibromyalgia, or if any were treated for anxiety and depression.

The valuations of profiles vary by how sets of preference weights for health profiles were developed. The estimated value for 22232 was -0.016 in the UK study [4] and -0.052 in the French study [11]; this particular health profile was not rated "worse than death" in 5 other published EQ-5D valuations (United States, Denmark, Spain and Japan) [15]. The valuation of this health profile as "worse than death" for UK and French weights may be an artifact of the method used to estimate valuations for health profiles not directly valued by the general population sample but extrapolated from the other health profiles evaluated. However, the same preference-elicitation

methodology was used in a Spanish study [16], and a reduced study design was used in the Japan study [17], for 17 hypothetical EQ-5D health states evaluated as compared with 42 in the UK study.

When assessed by clinical measures (eg, swollen joint counts, pain on a visual analog scale [VAS], functional disability), the health status of patients with a negative EQ-5D score was worse than for patients with EQ-5D ≥ 0 and these differences were important and probably clinically meaningful. Nevertheless at the individual level, variability of measured characteristics in patients with EQ-5D < 0 was large: pain at rest on a VAS from 0 to 100, fatigue from 20 to 100, PCS of the SF-36 from 12 to 48 and MCS of the SF-36 from 12 to 58, respectively. Moreover the DAS28 ranged from 3.2 to 8.8, so patients with low disease activity can also present a state "worse than death". This study demonstrated a dichotomy between "perceived health" and "clinical status."

Of the 24 patients with a negative EQ-5D score at 2 years, 7 had a negative score at 1 year and 12 at baseline. Only 3 patients had a negative score at all times, and these patients had very active disease at baseline. At 2 years, 2 of these 3 patients were in DAS28 remission but they had all high functional disability, fatigue and low mental health. So the EQ-5D < 0 reflected health state more than a "personality feature".

Utilities measure the patient preference for health state by population-assigned weights, whereas HRQoL scales assess the patient assessment of functioning and well-being. Although these 2 concepts are related, they are not identical [18], and measuring both may lead to a better understanding of health state in patients with a patient perspective and a mixed patient and societal perspective. According to the societal perspective, EA patients with a negative EQ-5D score should be a priority for health care: they have specific needs necessitating better management of pain and psychological state.

**Key messages:**

- High pain/discomfort was the key domain leading to states “worse than death” in early arthritis.
- High level of disability and mental distress are independent determinants of state “worse than death”.
- Early arthritis patients have specific needs necessitating better management of pain and psychological state.

**Acknowledgements**

We thank Nathalie Rincheval for expert study monitoring and data management; S. Martin (Paris Bichat) for measuring CRP level, IgA and IgM rheumatoid factor and anti-CCP antibodies; and all the investigators who recruited and followed the patients (F. BERENBAUM, Paris- Saint Antoine, MC. BOISSIER, Paris-Bobigny; A. CANTAGREL, Toulouse; B. COMBE, Montpellier, M. DOUGADOS, Paris-Cochin; P. FARDELLONE and P BOUMIER, Amiens; B. FAUTREL, P. BOURGEOIS, Paris-La Pitié; RM. FLIPO, Lille; P. GOUPILLE, Tours; F. LIOTE, Paris-Lariboisière; X. LE LOET, Rouen; X. MARIETTE, Paris Bicêtre; O. MEYER, Paris Bichat; A. SARAUX, Brest; T. SCHAEVERBEKE, Bordeaux; J. SIBILIA, Strasbourg).

**Conflict of interest statement**

The authors declare no conflict of interest.

**Funding statement**

"An unrestricted grant from Merck Sharp and Dohme (MSD) was allocated for the first 5 years of the ESPOIR cohort study. Two additional grants from INSERM were obtained to support part of the biological database. The French Society of Rheumatology, Abbott, and Wyeth-Pfizer also supported the ESPOIR cohort study."

**References**

[1] Harrison MJ, Bansback NJ, Marra CA, Drummond M, Tugwell PS, Boonen A. Valuing health for clinical and economic decisions: directions relevant for rheumatologists. J Rheumatol. 2011;38:1770-5.

[2] Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 1977; 296:716–21.

[3] Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Med Decis Making. 2005;25:460-7.

[4] Dolan P. Modeling valuations for EuroQol health states. Med.Care 1997;35:1095-1108.

[5] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J.Health Econ. 2002;21:271-292.

[6] Khanna D, Tsevat J. Health-related quality of life--an introduction. Am. J. Manag. Care 2007;13 Suppl 9:S218-23.

[7] Combe B, Benessiano J, Berenbaum F, et al. The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. Joint Bone Spine 2007;74:440-445.

[8] Prevoo ML, van 't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38:44-48.

[9] Fries JF, Spitz P, Kraines RG, et al. Measurement of patient outcome in arthritis. Arthritis Rheum.1980;23:137-45.

[10] Barton GR, Sach TH, Avery AJ, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged >or= 45 years. Health Econ. 2008;17:815-832.

[11] Chevalier J, de Pouvourville G.Valuing EQ-5D using Time Trade-Off in France. Eur J Health Econ. 2011 Sep 21. [Epub ahead of print]

[12] Macran S, Kind P. "Death" and the valuation of health-related quality of life. Med Care. 2001;39:217-27.

[13] De Wit GA, Busschbach JJV, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? Health Econ 2000;9:109–26.

[14] Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Farragher TM, Verstappen SM, Hassell A, Symmons DP. Why do patients with inflammatory arthritis often score states "worse than death" on the EQ-5D? An Investigation of the EQ-5D classification system. Value Health. 2009;12:1026-34.

[15] Adams R, Walsh C, Veale D, et al. Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics. 2010;28:477-87.

[16] Badia X, Roset M, Herdman M, Kind P.A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making. 2001;21:7-16.

[17] Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al.Estimating an EQ-5D population value set: the case of Japan. Health Econ. 2002;11:341-53.

[18] Apolone G, De Carli G, Brunetti M, Garattini S. Health-related quality of life (HR-QOL) and regulatory issues. An assessment of the EMEA recommendations on the use of HR-QOL measures in drug approval. Pharmacoeconomics 2001, 19:187-95.

**Table 1:** Characteristics of patients included in ESPOIR cohort at baseline (n=813) and by EQ-5D score

| Characteristic | **All patients (n=813)** Mean ± SD *Median [min-max]* | **Patients with EQ-5D < 0 (n=90)** Mean ± SD *Median [min-max]* | **Patients with EQ-5D ≥ 0 (n=718)** Mean ± SD *Median [min-max]* | **P value[a]** |
|---|---|---|---|---|
| Age (years) | 48.1 ± 12.6 | 47.8 ± 13.4 | 48.1 ± 12.5 | 0.80 |
| Female sex, no. (%) | 624 (76.7) | 70 (80.0) | 550 (76.6%) | 0.80 |
| >12 years of education, no. (%) | 255 (31.4) | 17 (18.9) | 238 (33.1) | **0.006** |
| Employed, no. (%) | 519 (64.2) | 53 (59.0) | 466 (64.9) | 0.26 |
| Married or living together, no. (%) | 594 (73.1) | 62 (68.9) | 528 (73.5) | 0.35 |
| DAS28** | 5.11 ± 1.31 | 6.34 ± 1.22 | 4.94 ± 1.23 | **<0.0001** |
| | *5.07 [1.45-8.80]* | *6.40 [3.20-8.80]* | *4.93 [1.45-8.52]* | |
| HAQ score | 0.98 ± 0.68 | 1.98 ± 0.44 | 0.85 ± 0.60 | **<0.0001** |
| | *0.88 [0-2.86]* | *2.00 [0.88-2.88]* | *0.88 [0-2.63]* | |
| SF-36 PCS | 37.4 ± 8.9 | 27.7 ± 6.2 | 38.5 ± 8.5 | **<0.0001** |
| | *37.5 [12.2-59.7]* | *26.5 [12.2-48.1]* | *38.4 [12.4-59.7]* | |
| SF-36 MCS | 39.5 ± 10.9 | 30.2 ± 8.0 | 40.7 ± 10.7 | **<0.0001** |
| | *39.4 [12.4-69.0]* | *29.0 [12.4-49.4]* | *40.3 [12.4-69.0]* | |
| Pain at rest (VAS, 0-100 mm) | 37.2 ± 27.6 | 50.4 ± 29.2 | 35.4 ± 26.9 | **<0.0001** |
| | *35.0 [0-100]* | *50.0 [0-100]* | *32.5 [0-100)* | |
| Pain during activities (VAS, 0=100 mm) | 54.8 ± 26.0 | 72.9 ± 22.6 | 52.5 ± 25.4 | **<0.0001** |
| | *56.0 [0-100]* | *78.0 [0-100]* | *54.0 [0-100]* | |
| Fatigue | 47.8 ± 28.2 | 65.1 ± 23.9 | 45.4 ± 28.0 | **<0.0001** |
| | *50.0 [0-100]* | *70.0 [21-100]* | *47.0 [0-100]* | |
| Erythrocyte sedimentation rate | 29.4 ± 24.6 | 46.7 ± 31.3 | 27.3 ± 22.8 | **<0.0001** |
| C-reactive protein level * | 20.3 ± 32.4 | 41.1 ± 43.0 | 17.7 ± 30.0 | **<0.0001** |

| | | | | |
|---|---|---|---|---|
| Rheumatoid factor*, no. (%) | 372 (45.8) | 48 (53.3) | 322 (44.8) | 0.13 |
| Anti-CCP2 antibodies*, no. (%) | 315 (38.8) | 46 (51.1) | 269 (37.5) | **0.012** |
| Van der Heijde-modified Sharp score $ | 5.7 ± 7.70 | 5.73 ± 7.33 | 5.76 ± 7.76 | 0.97 |
| EQ-5D score | 0.498 ± 0.316 | -0.103 ± 0.084 | 0.573 ± 0.246 | **<0.0001** |
| | *0.577 [-0.53;1]* | *-0.052* [-0.530;-0.011] | *0.052* [0.005;1.000] | |
| ACR/EULAR criteria for rheumatoid arthritis, no. (%) | 641 (78.8) | 80 (88.9) | 558 (77.8) | 0.015 |

[a] Comparison between patients with an EQ-5D<0 and EQ-5D ≥0 at baseline.

HAQ=Health Assessment Questionnaire; PCS=physical component score; MCS=mental component score; VAS=visual analog scale; ACR/EULAR=American College of Rheumatology/European League Against Rheumatism

*Baseline C-reactive protein level (normal < 10 mg/l); IgM and IgA rheumatoid factor (ELISA, Menarini, France; positive > 9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive > 50 U/ml) detected in all patients by the same technique in a central lab (Paris-Bichat).

**for 9.5% of patients, the Disease Activity Score in 28 joints (DAS28) was ≤3.2, for 44.6% 3.2–5.0 and for 45.9% >5.1

$ Van der Heijde-modified Sharp score: for 715 patients, 160 (22.3%) had erosions on hands and/or feet at baseline.

**Table 2:** Discriminant capacity of levels of 3 domains of the EQ-5D for patients with early arthritis and EQ-5D < 0

| **Outcomes** | **Mobility** | | | **P value *** |
|---|---|---|---|---|
| | **No problem** | **Some problems** | **Extreme problems** | |
| No. of patients | 2 | 83 | 5 | |
| SF-36 PCS | 33.2 | 26.4 | 30.8 | 0.15 |
| HAQ | 1.87 | 2.00 | 2.37 | 0.41 |
| | **Pain or discomfort** | | | |
| | **No problem** | **Some problems** | **Extreme problems** | |
| No. of patients | 0 | 8 | 82 | |
| Pain at rest (VAS, 0-100 mm) | _ | 33.0 | 50.0 | 0.18 |
| Pain during activities (VAS, 0-100 mm) | _ | 74.0 | 79.5 | 0.76 |
| | **Anxiety or depression** | | | |
| | **No problem** | **Some problems** | **Extreme problems** | |
| No. of patients | 1 | 52 | 35 | |
| SF-36 MCS | 25.0 | 31.5 | 27.7 | 0.005 |

Data are represented as median values

* Kruskal-Wallis test

HAQ=Health Assessment Questionnaire; PCS=physical component score; MCS=mental component score; VAS=visual analog scale

**Table 3:** Determinants of health state "worse than death" (EQ-5D < 0) at baseline and 6, 12 and 24 months by multivariate analysis for patients with early arthritis

| **Determinant** | **Baseline** | | | **6 months** | | | **12 months** | | | **24 months** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EQ-5D<0 n=90 | EQ-5D≥0 n=718 | OR [95% CI] | EQ-5D<0 n=25 | EQ-5D≥0 n=731 | OR [95% CI] | EQ-5D<0 n=23 | EQ-5D≥0 n=703 | OR [95% CI] | EQ-5D<0 n=24 | EQ-5D≥0 n=667 | OR [95% CI] |
| HAQ (0-3=worst) | 1.98±0.44 | 0.85±0.60 | **22.6 [10.6;48.0]** | 1.85±0.36 | 0.46±0.51 | **66.6 [15.2;291.6]** | 1.74±0.59 | 0.48±0.54 | **23.1 [7.3;73.5]** | 1.81±0.44 | 0.43±0.52 | **20.6 [5.7;74.2]** |
| MCS (0-100=best) | 30.2±8.0 | 40.7±10.7 | **0.92 [0.88;0.95]** | 30.4±7.0 | 44.8±10.8 | 0.94 [0.87;1.01] | 30.2±7.1 | 45.0±11.3 | **0.93 [0.88;0.99]** | 27.4±7.3 | 46.3±10.9 | **0.89 [0.83;0.96]** |
| Fatigue (0-100=worst) | 65.1±23.9 | 45.4±28.0 | 1.00 [0.99;1.01] | 69.7±22.7 | 37.8±27.9 | 1.02 [0.99;1.04] | 69.1±24.8 | 36.3±28.3 | 0.99 [0.97;1.02] | 75.2±22.2 | 33.2±27.8 | 1.02 [0.99;1.05] |
| DAS28 (0.49-9.07=worst) | 6.34±1.22 | 4.94±1.23 | 1.18 [0.89;1.56] | 4.91±0.96 | 3.34±1.38 | 0.98 [0.54;1.78] | 4.47±1.77 | 3.10±1.34 | 0.76 [0.47;1.24] | 4.81±2.1 | 2.84±1.30 | 1.12 [0.68;1.83] |
| Pain at rest (0-100=worst) | 50.4±29.2 | 35.4±26.9 | 0.99 [0.98;1.01] | 54.3±20.8 | 22.1±23.3 | 1.00 [0.98;1.03] | 53.1±29.0 | 21.1±23.7 | 1.02 [0.99;1.04] | 54.7±30.0 | 18.2±22.2 | 1.00 [0.97;1.02] |
| CRP level or ESR* | 41.1±43.0 | 17.7±30.0 | 1.00 [0.99;1.01] | 23.2±20.8 | 16.0±13.7 | 1.00 [0.97;1.04] | 22.8±29.6 | 15.0±13.0 | 0.98 [0.97;1.02] | 24.6±21.2 | 15.3±13.6 | 1.01 [0.97;1.06] |
| Educational level # no. (%) | 17 (18.9) | 238 (33.1) | 1.20 [0.53;2.51] | 1 (4.0) | 243 (33.2) | 5.77 [0.55;60.4] | 3 (13.0) | 231 (32.9) | 1.06 [0.25;4.50] | 5 (20.9) | 223 (33.4) | 0.46 [0.12;1.76] |

Data are mean±SD unless indicated
OR=odds ratio; 95% CI=95% confidence interval; MCS=Mental component of the SF-36; HAQ=Health Assessment Questionnaire; DAS28=Disease Activity Score in 28 joints
* Baseline-centralised C-reactive protein (CRP) level was available only at baseline; erythrocyte sedimentation rate (ESR) was available at 6, 12 and 24 months.
# Educational level was defined by >12 years: yes/no.
Significant results are in bold.

**Figure 1:** Frequency distribution of patients with early arthritis and health state "worse than death" (EQ-5D < 0) at baseline (n=90).

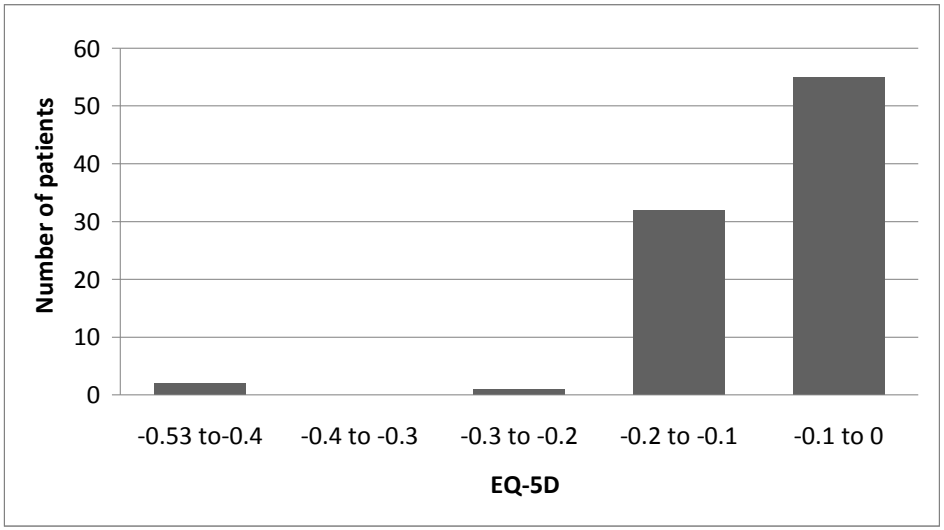


**Figure 2:** Distribution of levels in each domain of the EQ-5D for patients with early arthritis and health state "worse than death" (EQ-5D < 0) at baseline

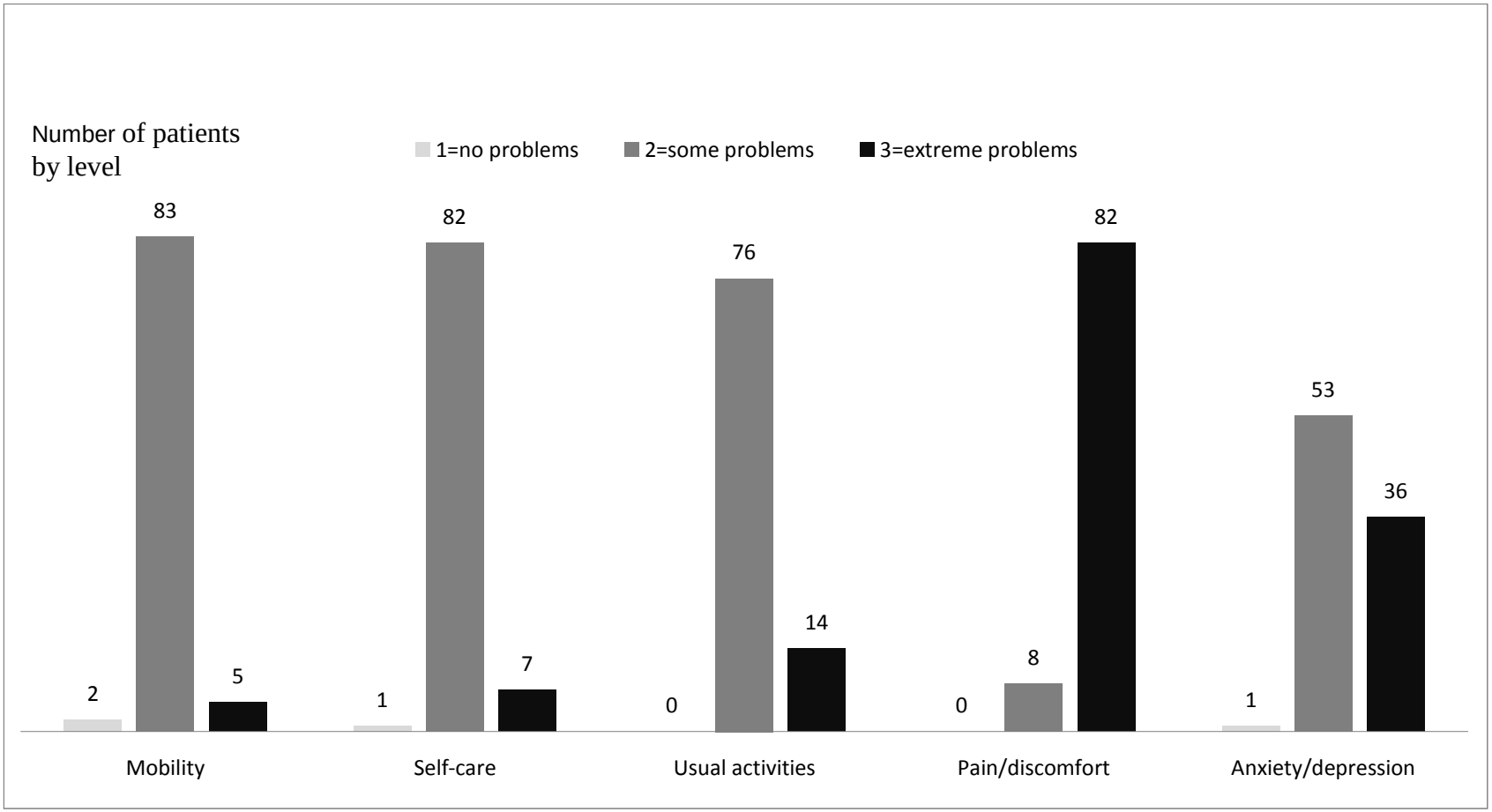
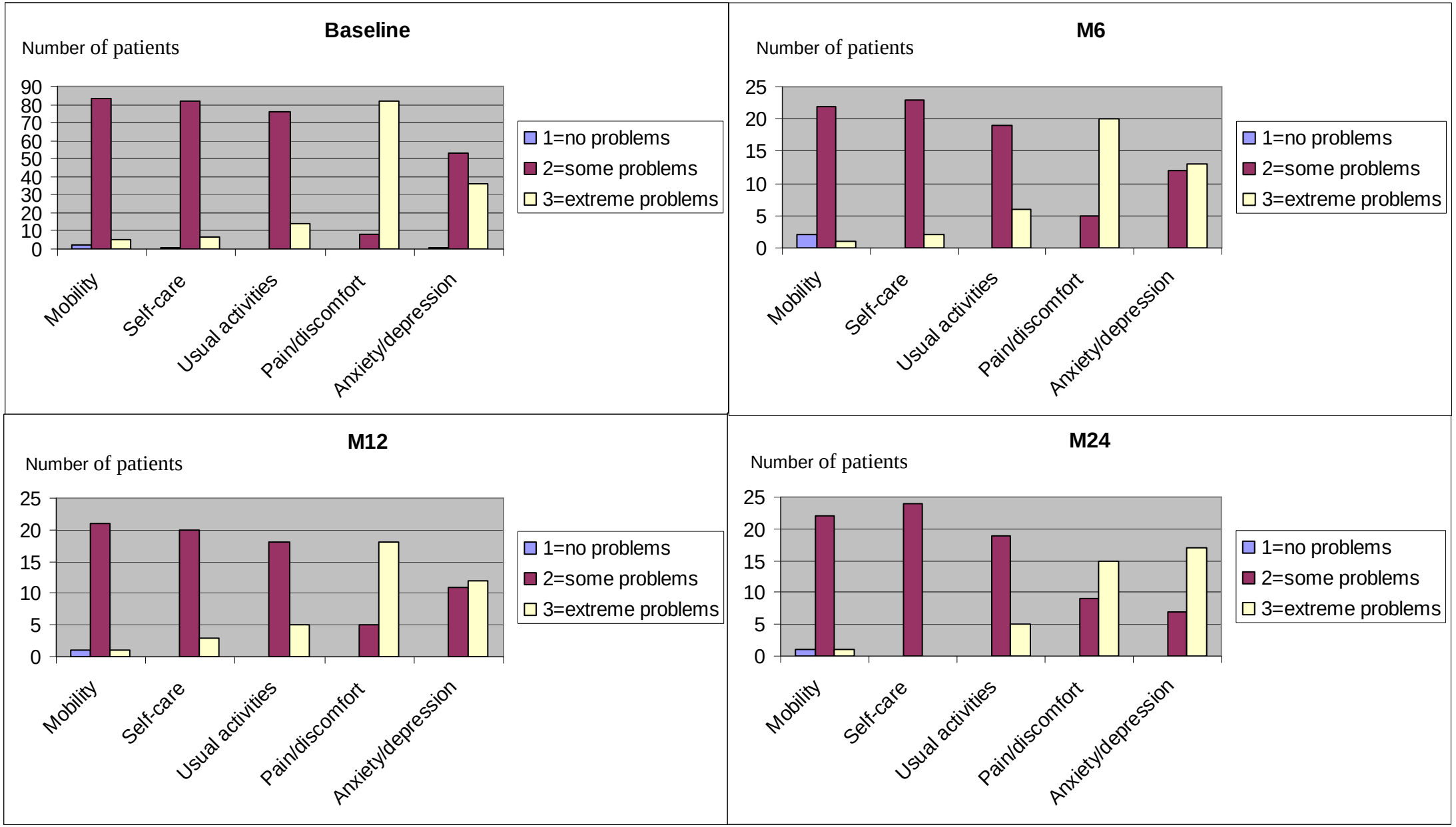


**Figure 3:** Distribution of levels in each EQ-5D domain over 2 years for patients with early arthritis and health state "worse than death" (EQ-5D < 0)



# The impact of functional disability on EQ-5D and SF-6D utility discrepancies in early arthritis: Results from the ESPOIR cohort

C Gaujoux-Viala[1,2], B Fautrel[2], K Hosseini[1,3], RM Flipo[4], F Guillemin[1,3], AC Rat[1,3]

[1] Lorraine University, Paris Descartes University, EA 4360 Apemac, Nancy, France

[2] Paris 6 – Pierre et Marie Curie University; AP-HP, Department of Rheumatology, Pitié-Salpêtrière Hospital- GRC-UPMC 08 – EEMOIS, Paris, France

[3] INSERM, CIC-EC CIE6, Nancy, France

[4] Department of Rheumatology, Lille University 2, France

**Key words**: utility, SF-6D, EQ-5D, early arthritis

**Institution to which the work should be attributed**: Lorraine University, Paris Descartes University, EA 4360 Apemac; Paris 6 – Pierre et Marie Curie University; Department of Rheumatology, Pitié-Salpêtrière Hospital, Paris, France

GRC-UPMC 08 – EEMOIS

**Authors:**

Cécile Gaujoux-Viala, MD, Bruno Fautrel, MD, PhD, MD, PhD, Kossar Hosseini, MD, Francis Guillemin, MD, PhD, René-Marc Flipo, MD, PhD, Anne-Christine Rat, MD, PhD

**Corresponding author**: Dr Cécile Gaujoux-Viala, corresponding author

Université Paris 6 – Pierre & Marie Curie

AP-HP, Groupe hospitalier Pitié-Salpêtrière,

service de Rhumatologie,

83 boulevard de l'Hôpital 75651 Paris cedex 13, France

Phone: +33 1.421.77821 – Fax : +33.1.421.77955

Email: cecilegaujouxviala@yahoo.fr

**Conflict of interest :** none

**Short running footline**: Utility in early arthritis

**Word count:** 2851

**Abstract**. 247 words

**Objectives**: To assess characteristics associated with a substantial difference between 2 indirect utility measures, the EQ-5D and SF-6D, in early arthritis (EA).

**Methods**: Multivariate linear regression was used to analyze data for 813 EA patients to determine the EA features independently associated with EQ-5D and SF-6D scores and difference in scores. Polytomous logistic regression was used to analyze 3 categories as dependent variables: │SF-6D – EQ-5D│>0.074 (Minimal Important Difference) and SF-6D>EQ-5D, │SF-6D – EQ-5D│>0.074 and EQ-5D>SF-6D, and no substantial difference.

**Results**: High HAQ score (high disability), pain and low mental status (AIMS2-SF score) were significant determinants of low EQ-5D score. The HAQ score explained 41.4% of the variance. High HAQ score, pain, patient global assessment and fatigue, and low mental status were associated with low SF-6D score. The HAQ score explained 49.3% of the variance. Most patients (66.5%) showed │SF-6D – EQ-5D│ > 0.074 (mean difference 0.064 [range –0.421 to 0.941]). For these patients, high disability and CRP level and low mental status were associated with the difference between scores. The HAQ score explained 43.3% of the variance. With polytomous logistic regression, HAQ score, high pain and low mental status were associated with a substantial difference between the utility scores. The influence of the HAQ was bidirectional: high disability was associated with probability of the SF-6D superiority, and low disability was associated with the probability of the EQ-5D superiority.

**Conclusions**: Functional disability is the key element in the substantial difference between the EQ-5D and SF-6D utility measures in EA patients.

**Significance and innovation**:

- This study is the first to examine in a real life setting the determinants of the EQ-5D and SF-6D utility measures and the difference between the 2 in patients with early arthritis: most of the 813 patients examined showed a substantial difference between the 2 utility scores.
- Functional disability (the Health Assessment Questionnaire score) explained most of the difference between the scores.
- The effect of functional disability was bidirectional: high disability was associated with probability of the SF-6D score being greater than the EQ-5D score, and low disability was associated with the probability of the EQ-5D being greater than the SF-6D score.
- EQ-5D and SF-6D are not interchangeable, especially in patients with low functional ability, and the results of cost-utility studies involving different utility instruments must not be compared.

## Introduction

The utility concept was developed by health economists. Assessment of utility assigns a numeric value from 0 to 1 for health states, 0 indicating death and 1 a state of perfect health. The values reflect the preference for a health state in a situation of choice that includes uncertainty (in terms benefits or risks) or sacrifice (eg, life years). Methods such as standard gamble [SG] and time trade- off [TTO] may be used to measure health states directly, but they are less suitable for clinical research and less widely used for feasibility reasons. Therefore, indirect utility assessment tools such as the EQ-5D [1] and SF-6D [2] have been developed.

The EQ-5D and the SF-6D are indirect preference-based instruments of health-related quality of life being increasingly used for economic evaluation of clinical interventions and health programs. The indirect health utility assessments involve use of population-assigned weights to calculate utility scores for particular health states with multidomain health-status questionnaires completed by patients [3]. Utility values are used to calculate quality-adjusted life years (QALYs), which are outcome measures that combine quantity and quality of life [4]. QALYs are the number of years of life in which a person is in a perfect health state. They are considered equivalent to years actually lived in a given health state multiplied by the utility value of this state [4]. QALYs are used to assess the benefit of the health interventions in economic evaluations, such as cost-utility analysis. They are particularly informative for health policy decisions because they allow for direct comparison of the efficiency of health care resource expenditure across a wide range of conditions and treatments.

Although the theoretical concept of utility implies that one specific health state has one utility score, regardless of how it is measured, different instruments can give different scores and so different QALYs estimates [5]. Several studies have compared EQ-5D and SF-6D scores for patients with a particular clinical condition; a common finding is small but clinically significant differences between the utility estimates [6-9]. However, few comparisons exist for early arthritis (EA). In a previous study [10], we demonstrated systematic disagreement

between EQ-5D and SF-6D scores for EA, especially for patients with worse clinical outcomes: the utility score was systematically lower with the EQ-5D, possibly because the 2 tools differently reflect the multiple facets of quality of life or utility.

Here, in order to better understand what drive their differences, we aimed to assess, in a large group of patients with EA, 1) the socio-demographic and clinical characteristics at baseline associated with EQ-5D and SF-6D scores and 2) which of these variables are associated with substantial difference between the scores.

## Material and Methods

### *Patients*

Between December 2002 and March 2005, 813 patients with EA from 14 French regional centers were included in the ESPOIR cohort [11]. Inclusion criteria were age 18 to 70 years, more than 2 swollen joints for >6 weeks and <6 months, suspected or confirmed diagnosis of RA, and not taking disease-modifying antirheumatic drugs or steroids (except if < 2 weeks). Patients were excluded if the referring physician considered that they had other clearly defined inflammatory rheumatic diseases. At baseline, data for a set of clinical and biological variables were recorded, by the Disease Activity Score for 28 Joints (DAS28), a composite index of disease activity (range: 0.49 -9.07 : with 9.07 high activity) [12], a functional ability questionnaire, the Health Assessment Questionnaire (HAQ) (range: 0-3, with 3 the worst incapacity) [13], and pain at rest on a 100-mm visual analogue scale (range: 0-100, with 100 the worst pain). Fatigue severity over the previous 48 hours was assessed by a 100-mm visual analog scale, (range: 0-100, with 100 the worst fatigue). Furthermore, patients completed other self-administered patient-reported outcome measures, including health-related quality of life questionnaires, the Arthritis Impact Measurement Scale 2–Short Form (AIMS2-SF) [14], the EQ-5D and the SF-36 [15]. Baseline C-reactive protein level, IgM and IgA rheumatoid factor (ELISA, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin,

France; positive > 50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

The protocol of the ESPOIR Cohort study was approved by the ethics committee of Montpellier, France. All patients gave their signed informed consent before inclusion.

### *Utility measures*

#### *EQ-5D*

The EQ-5D consists of 5 items or dimensions (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression) with 3 answer modes: no problem (1), some problems (2), or severe problems (3). The combination of the 5 dimensions with the 3 answer modes produces 243 possible health states described as vectors ranging from 11111 (no problems in any domain) to 33333 (severe problems in all five domains). The EQ-5D can be reported in terms of a 5-digit profile indicating the level on each dimension or a preference-based single index number. The latter is obtained by applying algorithms that link the 5-digit health state description to mean values for members of the general population in the UK using the TTO method [1]. This algorithm yields a range from -0.59 to +1, with 1.00 representing full health, 0 death, and negative value, a health state labelled “worse than death” [1].

#### *SF-6D*

The SF-6D, developed from 11 questions of the SF-36, is a multidimensional health classification system assessing the 6 health domains of physical functioning: role limitation, social functioning, pain, mental health, and vitality, with 4 to 6 levels for each domain. An SF-6D health state is defined by 1 level from each domain, which results in a total of 18,000 possible health states. The SF-6D scoring algorithm was developed with the SG method and a sample of 249 SF-6D health states in a representative sample of the UK population [2]. Utility

scores generated by the SF-6D range from 0.29 to 1.00, with 1.00 representing full health and 0.29 the worst possible health state (ie, all domains being at the worst level).

### *Statistical analysis*

#### *Global utility scores*

Descriptive statistics (mean [SD], median, minimum, maximum) and EQ-5D and SF-6D utility scores at baseline were computed. Variables potentially associated with the EQ-5D and SF-6D were analysed by bivariate analysis. Variable with values significant at $p \leq 0.20$ were included in multivariate linear regression analysis (stepwise selection) to assess which baseline variables are associated with the EQ-5D and SF-6D scores in EA. Partial R-squares were used to assess the relative importance of variables to the variation in EQ-5D and SF-6D scores.

Due to the bimodal distribution of the EQ-5D score with a cut-off at 0.3, multivariate logistic regression was used to determine specific aspects of EA independently associated with EQ-5D < 0.3 at baseline. We compared the characteristics and health status of patients with EQ-5D < 0.3 and EQ-5D ≥ 0.3 at baseline: qualitative variables were compared by chi-square test or Fisher's exact test as appropriate and quantitative variables with one-way ANOVA or Mann–Whitney U-test as appropriate. A logistic regression model was then constructed with the dependent variable EQ-5D < 0.3 (yes/no) at baseline. The model entered candidate variables with a p value <0.2 on bivariate analysis with a stepwise selection. If several outcomes had similar underlying constructs, only one was selected to avoid collinearity.

#### *Difference between the SF-6D and the EQ-5D*

Descriptive statistics (mean [SD], median, minimum, maximum) and distributions of the difference between the 2 utility measures (i.e., SF-6D – EQ-5D) at baseline were computed.

To determine which variables are associated with a substantial difference between the 2 instruments, we chose a cut-off of 0.074 between the scores on the basis of the higher estimate

of the minimal important difference (MID) using anchor-based methods published for the EQ-5D and SF-6D [16]. Walters et al. compared the MID for the SF-6D and EQ-5D from 8 longitudinal studies that used both instruments with 11 patient groups. The mean MID for the SF-6D was 0.041 and for the EQ-5D, 0.074 [16]. Thus, we considered that patients with $|$SF-6D – EQ-5D$|$ > 0.074 showed a substantial difference between the 2 scores. In these patients, variables potentially associated with the difference between the 2 utility scores at baseline were analysed by bivariate analysis. Variables with values significant at $p \leq 0.20$ were included in a multivariate linear regression (stepwise selection). Partial R-squares were used to explain the relative importance of individual variables to the substantial difference between the SF-6D and EQ-5D scores.

To assess the determinants of a substantial utility difference, patients were divided into 3 categories: $|$SF-6D – EQ-5D$|$>0.074 and SF-6D > EQ-5D (category -1); $|$SF-6D – EQ-5D$|$> 0.074 and EQ-5D > SF-6D (category 1); and others, corresponding to no substantial difference (category 0) as the reference category. Variables collected at baseline were analysed by polytomous logistic regression, with the 3 categories as dependent variables. The models provided regression coefficients for independent variables, antilogs of which were odds ratios (ORs) expressing the effect, when changing one unit of the independent variable, on the odds of being a patient with substantial difference and EQ-5D > SF-6D (category 1) or a patient with substantial difference and SF-6D > EQ-5D (category -1) as compared with a patient with no substantial difference (category 0), holding other variables constant in the equation model. A stepwise regression model entered candidate variables ($p < 0.2$ on bivariate analysis).

Due to heterogeneity in potential MIDs, we performed a sensitivity analysis with a cut-off for a substantial difference of 0.129 on the basis of the higher estimate of the MID published for the EQ-5D and SF-6D in early RA [16]. We also use a cut-off of 0.06: the MID for SF-6D was estimate around 0.03 in several studies [16-17] and the EQ-5D scale having

approximately twice the range of the SF-6D scale so the estimate of the MID for the EQ-5D is almost double that of the estimate of the MID for the SF-6D [16].

Because the lower bounds of the 2 instruments differ (EQ-5D allows for negative utility values and the SF-6D score cannot be < 0.296), we performed a sensitivity analysis. EQ-5D and SF-6D utility scores were transformed linearly to fit the range 0-1 to retain scale proportionality and the cut-off for a substantial difference was accordingly modified.

We also performed a sensitivity analysis by using AIMS2-SF physical score instead of HAQ.

All analyses involved use of SAS v9.2 (SAS Inst., Cary, NC, USA). A $p<0.05$ was considered statistically significant.

## Results

### *Characteristics of the population*

Table 1 shows the demographic and clinical characteristics of the 813 patients in the ESPOIR cohort at inclusion. In total, 641 (78.8%) patients fulfilled the 2010 American College of Rheumatology–European League Against Rheumatism criteria for RA.

### *Global utility scores*

#### *EQ-5D*

The distribution of EQ-5D scores was bimodal (Figure 1). At baseline, the mean score was 0.52 ± 0.31, median 0.66 [range –0.59 to 1]. Few missing values were observed: 0.6%.

By multivariate linear regression, high HAQ (0-3: the worst), high pain and low mental status (AIMS2-SF) (0-100: the best) were significant determinants of a low EQ-5D at baseline. HAQ and AIMS2-SF-mental scores explained 41.4 and 7.9% of the variance respectively (table 2).

Patients with EQ-5D < 0.3 had lower educational level, greater disease activity (DAS28), higher HAQ score (greater functional disability), lower mental score, more pain at rest and fatigue, and higher C-reactive protein level than patients with EQ-5D ≥ 0.3 (all $p<0.001$) (Table 2). We included level of education, DAS28, HAQ score, mental status (AIMS2-SF), pain at rest, fatigue, and CRP level in the multivariate logistic regression model. Increased HAQ score, pain, low mental status and low educational level were independently associated with EQ-5D score <0.3 at baseline (Table 3).

#### *SF-6D*

The distribution of SF-6D scores was near normal (Figure 1). At baseline, the mean score was 0.58 ± 0.11; median 0.58 [range 0.30 to 0.92]. Few missing values were observed: 1.2%.

By multivariate linear regression, high HAQ score, pain at rest, patient global evaluation, and fatigue and low mental status and educational level were significantly associated with low SF-6D score. HAQ and AIMS2-SF mental health scores explained 49.3% and 14.5% of the variance, respectively (Table 2).

*Difference between the SF-6D and EQ-5D*

The distribution of the difference between the 2 utility scores was bimodal (Figure 2). The mean difference in utility scores for the 2 measures was 0.064 [range -0.421 to 0.941]). Most patients (66.5%) showed $|\text{SF-6D} - \text{EQ-5D}| > 0.074$ (MID). For these patients, the distribution of the difference between the 2 utility scores was near normal. High HAQ score and C-reactive protein level and low mental status explained the difference between the 2 scores. HAQ and AIMS2-SF mental health scores explained 43.3% and 2% of the variance, respectively (Table 2).

By polytomous logistic regression, HAQ score, high pain and low mental status were associated with a substantial difference between the EQ-5D and SF-6D scores. Low HAQ score and mental status, high pain and patient global assessment (representing worse condition) were associated with the probability of EQ-5D being substantially greater than the SF-6D as compared with no substantial difference. High HAQ score and pain and low mental status were associated with the probability of SF-6D being substantially greater than the EQ-5D (Table 4).

Due to heterogeneity in potential MIDs, we performed a sensitivity analysis with cut-off for a substantial difference of 0.129 and 0.06. Results were similar except for the polytomous regression with the cut-off of 0.129 where 2 additional variables became significant: high disease activity (DAS 28) was associated with the probability of EQ-5D being substantially greater than the SF-6D as compared with no substantial difference, and low educational level

(<12 years of education) was associated with the probability of SF-6D being substantially greater than the EQ-5D.

Given the concern that MID might be adversely affected by scaling differences between the EQ-5D and SF-6D, we performed a sensitivity analysis by a new polytomous logistic regression analysis with a recalculated MID and standardized EQ-5D and SF-6D. Interestingly, on sensitivity analysis, for 157 patients, the standardized EQ-5D was substantially greater than the standardized SF-6D, and 625 patients showed no difference in scores. None showed a standardized SF-6D substantially greater than the standardized EQ-5D. By logistic regression, only low mental status was associated with the probability that the standardized EQ-5D was substantially greater than the standardized SF-6D**.** By multivariate linear regression, only high HAQ score was significantly associated with a difference between the standardized utility scores, but the model only explained 0.82% of the variance.

Using AIMS2-SF physical score instead of HAQ score, results were similar except for the SF-6D with an inversion in the main determinants: the mental status and the physical status explaining 44.1% and 11.6% of the SF-6D variance, respectively.

## Discussion

This study is the first to examine in a real life setting the determinants of the utility measures EQ-5D and SF-6D and the difference between them in patients with EA. Most of the 813 patients examined showed a substantial difference between the 2 utility scores (>0.074, MID for the EQ-5D and almost twice the MID for the SF-6D). Functional ability and to a lesser extent, mental health status and pain had consequences for utility evaluation in EA patients. Mainly functional disability (HAQ score) explained the difference between the EQ-5D and SF-6D scores. The HAQ score explained 49% of the variance in the SF-6D score, 41% of the variance in the EQ-5D score, and up to 43% of the variance with substantial difference between the 2 scores.

Moreover, the influence of the HAQ was bidirectional: high disability (ie, high HAQ score) was associated with the probability that the SF-6D score was substantially greater than the EQ-5D score and low disability was associated with the probability that the EQ-5D was greater than the SF-6D. In contrast, low mental health score and high pain were associated with substantial difference between both scores, whether the EQ-5D or the SF-6D was the highest. This result was exactly the same with the use of AIMS2-SF physical score instead of HAQ score.

In a precedent study, we have shown that although the correlation between the 2 utility scores, the EQ-5D and SF-6D, was high, descriptive statistics revealed systematic disagreement at both the low and high ends of the utility scales. In particular, EQ-5D values <0.5 corresponded to markedly high SF-6D scores. In addition, a wide range of SF-6D scores (0.58–0.85) was associated with an EQ-5D score of 1.0. Bland-Altman plots also displayed lack of agreement between the 2 measures, particularly at the low end of the utility scales. The lack of agreement increased with increasing disease activity and incapacity [10].

Several reasons might explain the differences between the utility scores. First, the 2 systems have different operational definitions of domains and functional levels within each domain. As well, EQ-5D utility scores are TTO based, which tends to result in high values for mild states, whereas SF-6D scores are SG based, which tends to result in high values for severe states [18,19]. Moreover, the SF-6D may be more sensitive for patients experiencing mild to moderate health problems because of its larger descriptive system.

Second, the health descriptive system of the SF-6D does not allow for negative values and so assigns a 0.296 value to the most severe health state, whereas the EQ-5D score allows for negative scores. We previously found that EA patients with an EQ-5D score “worse than death” (n=90) had high scores for pain and physical function, and a large proportion of these patients had the maximum scores for the pain dimension and moderate scores for all other dimensions [20]. These results confirmed those of another study investigating the health states

of patients with inflammatory arthritis with an EQ-5D score “worse than death” [21]. We also concluded that disability (high HAQ score) and low mental state (SF-36 mental component score) were always independently associated with a state “worse than death”.

Finally, we found low EQ-5D scores for patients with severe disability. This result may be explained by the content of EQ-5D. Of the 5 dimensions, 4 (mobility, self-care, usual activity and pain/discomfort) are likely to be particularly affected for patients with EA. A study comparing EQ-5D and SF-6D in 7 diseases [22] showed higher mean differences between the 2 instruments in osteoarthritis than in diseases focusing on pain and discomfort such as irritable bowel syndrome.

Our study has some limitations. The scoring algorithms used for both instruments were developed from data for a general population in the United Kingdom because no such algorithm was available in France for the SF-6D at the time of the study. Such an algorithm for the EQ-5D has been developed in France, but we chose to use algorithms for the same population for both the EQ-5D and SF-6D because comparisons might be more valid. However, we found the same association with disability and mental status when using the French value set of the EQ-5D that was derived from a population survey in France with 10-year TTOs [23]. The EQ-5D modelisation by linear regression was hampered because the EQ-5D scores showed a high level of skewness as compared with the normal distribution of the SF-6D scores but results were robust as checked by the use of White's robust variance-covariance matrix. We also used logistic regression and we found the same determinants (with educational level in addition). To take into account the heterogeneity in potential MIDs (especially for EQ-5D), we performed a sensitivity analysis with different cut-off for a substantial difference.

One of the strengths of our study is its investigation of a large, diverse group of patients with EA. The observational ESPOIR cohort aimed to include all patients with EA regardless of disease level, age and sex, and our study represents the performance of the instruments in a real-life setting.

This study aimed to develop a fuller understanding of the reasons for the difference between the SF-6D and EQ-5D utility scores in EA. Functional disability was the key element explaining a substantial difference between the 2 measures in EA patients. The EQ-5D and SF-6D scores are not interchangeable, especially in patients with worse functional ability, and the results of cost-utility studies involving different utility instruments must not be compared. It could be of interest that cost-utility analyses include both EQ-5D and SF-6D in sensitivity analyses.

**Acknowledgement**

An unrestricted grant from Merck Sharp and Dohme (MSD) was allocated for the first 5 years of the ESPOIR cohort study. Two additional grants from INSERM were obtained to support part of the biological database. The French Society of Rheumatology, Abbott, and Wyeth-Pfizer also supported the ESPOIR cohort study.

We thank Nathalie Rincheval for expert monitoring and data management and all the investigators who recruited and followed patients (F. BERENBAUM, Paris-Saint Antoine, MC. BOISSIER, Paris-Bobigny, A. CANTAGREL, Toulouse, B. COMBE, Montpellier, M. DOUGADOS, Paris-Cochin, P FARDELONNE, Amiens, B. FAUTREL, P BOURGEOIS, Paris-La Pitié, RM. FLIPO, Lille, Ph. GOUPILLE, Tours, F. LIOTE, Paris-Lariboisière, X. LE LOET Rouen, X MARIETTE, Paris Bicêtre, O MEYER, Paris Bichat, SARAUX, Brest, Th SCHAEVERBEKE, Bordeaux, J. SIBILIA, Strasbourg).

## References


[1] Dolan P. Modeling valuations for EuroQol health states. Med.Care 1997;35:1095-1108.

[2] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J.Health Econ. 2002;21:271-292.

[3] Khanna D, Tsevat J. Health-related quality of life--an introduction. Am. J. Manag. Care 2007;13 Suppl 9:S218-23.

[4] Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Med Decis Making. 2005;25:460-7.

[5] Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life years by different preference-based instruments. Med.Care 2003;41:791-801.

[6] Lillegraven S, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. Ann Rheum Dis. 2010;69:1762-7.

[7] Xie F, Li SC, Luo N, Lo NN, Yeo SJ, Yang KY, et al. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. Arthritis Rheum. 2007;57:1043-9.

[8] Marra CA, Esdaile JM, Guh D, Kopec JA, Brazier JE, Koehler BE, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. Med.Care 2004;42:1125-1131.

[9] Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whynes DK, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged >or= 45 years. Health Econ. 2008;17:815-832.

[10] Gaujoux-Viala C, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, et al. Comparison of the EQ-5D and the Short Form 6D utility measures in French patients

with early arthritis: Results from the ESPOIR cohort. J Rheumatol. 2011 Aug;38(8):1576-84.

[11] Combe B, Benessiano J, Berenbaum F, Cantagrel A, Daures JP, Dougados M, et al. The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. Joint Bone Spine 2007;74:440-445.

[12] Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44-48.

[13] Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum.1980;23:137-45.

[14] Guillemin F, Coste J, Pouchot J, Ghezail M, Bregeon C, Sany J, and the French Quality of Life in Rheumatology Group. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. Arthritis Rheum 1997;40:1267–74.

[15] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). Conceptual framework and item selection. Med Care. 1992 Jun;30(6):473-83.

[16] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005 Aug;14(6):1523-32.

[17] Gandhi PK, Ried LD, Bibbey A, Huang IC. SF-6D utility index as measure of minimally important difference in health status change. J Am Pharm Assoc (2003). 2012 Jan-Feb;52(1):34-42.

[18] Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. J.Health Econ. 1996;15:209-231.

[19] Tsuchiya A, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. J.Health Econ. 2006;25:334-346.

[20] Gaujoux-Viala C, Fautrel B, Guillemin F, Flipo R-M, Bourgeois P, Rat A-C. Who are the patients with early arthritis with “worse than death” scores on the EQ-5D? Results from the ESPOIR cohort. Rheumatology 2012 [Epub ahead of print]

[21] Harrison MJ, Davies LM, Bansback NJ, McCoy MJ, Farragher TM, Verstappen SM, et al. Why do patients with inflammatory arthritis often score states "worse than death" on the EQ-5D? An Investigation of the EQ-5D classification system. Value Health. 2009;12:1026-34.

[22] Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004;13:873-884.

[23] Chevalier J, de Pouvourville G.Valuing EQ-5D using Time Trade-Off in France. Eur J Health Econ. 2011 Sep 21. [Epub ahead of print]

**Table 1:** Characteristics of patients with early arthritis included in ESPOIR cohort at baseline

| **Characteristics** | **All patients (n=813)** |
|---|---|
| Age (years) | 48.1 ± 12.6 |
| Female sex, N (%) | 624 (76.7) |
| >12 years of education, N (%) | 255 (31.4) |
| Employed, N (%) | 519 (64.2) |
| Married or living together, N (%) | 594 (73.1) |
| DAS28 score (0.49 -9.07 : with 9.07 high activity) | 5.11 ± 1.31 |
| HAQ score (0-3, with 3 high disability) | 0.979 ± 0.684 |
| AIMS2-SF mental score (0-100, with 100 the best mental status) | 58.3 ± 19.9 |
| Patient global assessment (0-100, with 100 the worst health) | 59.8 ± 25.6 |
| Pain at rest (0-100, with 100 high pain) | 37.2 ± 27.6 |
| Fatigue (0-100, with 100 the worst fatigue) | 47.8 ± 28.2 |
| Erythrocyte sedimentation rate | 29.4 ± 24.6 |
| C-reactive protein level * | 20.3 ± 32.4 |
| Rheumatoid factor*, N (%) | 372 (45.8) |
| Anti-CCP2 antibodies*, N (%) | 315 (38.8) |
| Van der Heijde modified Sharp score $ | 5.7 ± 7.70 |
| SF-6D score | 0.58 ± 0.11 |
| EQ-5D score | 0.52 ± 0.31 |
| 2010 ACR-EULAR criteria for RA, N (%) | 641 (78.8) |

Data are mean±SD

*Baseline C-reactive protein level (normally < 10 mg/l); IgM and IgA rheumatoid factor (ELISA, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive > 50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

$ Van der Heijde modified Sharp score for 715 patients, 160 (22.3%) had erosions on hands and/or feet at baseline.

DAS28: disease activity score in 28 joints; HAQ: Health Assessment Questionnaire; AIMS2-SF: Arthritis Impact Measurement Scale 2–Short Form; ACR-EULAR: Americal College of Rheumatology–European League Against Rheumatism

**Table 2:** Multivariate linear regression analysis of determinants of the SF-6D and EQ-5D utility scores and the difference between the 2 scores at baseline

| Model | Variables | Beta | p | Partial $R^2$ |
|---|---|---|---|---|
| **EQ-5D** | HAQ | -0.20 | <0.0001 | 41.4% |
| $R^2$=0.51 | (0-3, with 3 the worst disability) | | | |
| | AIMS2-SF mental score | 0.005 | <0.0001 | 7.9% |
| | (0-100, with 100 the best mental status) | | | |
| | Pain at rest | -0.001 | 0.0009 | 0.8% |
| | (0-100, with 100 high pain) | | | |
| **SF-6D** | HAQ | -0.072 | <0.0001 | 49.3% |
| $R^2$=0.66 | AIMS2-SF-mental score | 0.002 | <0.0001 | 14.5% |
| | Patient global assessment | -0.0005 | <0.0001 | 1.8% |
| | (0-100, with 100 the worst health) | | | |
| | Pain at rest | -0.00025 | 0.0005 | 0.5% |
| | Fatigue | -0.0002 | 0.0186 | 0.2% |
| | (0-100, with the worst fatigue) | | | |
| | Educational level | 0.011 | 0.035 | 0.2% |
| **SF-6D – EQ-5D** | HAQ | 0.21 | <0.0001 | 43.3% |
| **in patients with** | AIMS2-SF mental score | -0.002 | <0.0001 | 2% |
| **│SF-6D – EQ5D│>0.074** | logCRP level | 0.036 | 0.0005 | 1.4% |
| $R^2$=0.47 | | | | |

**Table 3:** Determinants of EQ-5D < 0.3 at baseline by multivariate analysis for patients with early arthritis

| **Determinants** | **EQ-5D<0.3**<br>**n= 213** | **EQ-5D≥0.3**<br>**n=595** | **OR**<br>**[95% CI]** |
|---|---|---|---|
| HAQ<br>(0-3=worst) | 1.52 ± 0.67 | 0.78 ± 0.57 | 3.74 [2.69;5.21] |
| AIMS2-SF mental score<br>(0-100=best) | 43.4 ± 17.9 | 63.7±17.7 | 0.95 [0.94;0.96] |
| Pain at rest<br>(0-100=worst) | 49.1 ± 29.0 | 32.8 ± 25.7 | 1.01 [1.002;1.016] |
| Educational level #<br>no. (%) | 172 (80.7) | 381 (64.0) | 1.81 [1.15;2.87] |

Data are mean±SD unless indicated
OR=odds ratio; 95% CI=95% confidence interval;
# <12 years of education

**Table 4:** Multivariate polytomous logistic regression of determinants of the difference between the 2 utility measures (SF-6D – EQ-5D) at baseline

| Variable | Odds Ratio Estimate | 95% CI |
|---|---|---|
| | \|SF-6D – EQ-5D\| > 0.074 and SF-6D > EQ-5D (category -1, n=248), compared to no substantial difference (category 0, n=255) | |
| **HAQ** (0-3, with 3 high disability) | 2.184 | 1.537–3.101 |
| **AIMS2-SF mental score** (0-100, with 100 the best mental status) | 0.970 | 0.960–0.980 |
| **Pain** (0-100, with 100 high pain) | 1.015 | 1.007–1.023 |
| **Patient global assessment** (0-100, with 100 the worst health) | 1.006 | 0.997–1.015 |
| | \|SF-6D – EQ-5D\| > 0.074 and EQ-5D > SF-6D (category 1, n=279), compared to no substantial difference (category 0, n=255) | |
| **HAQ** | 0.585 | 0.419–0.817 |
| **AIMS2-SF mental score** | 0.987 | 0.977–0.997 |
| **Pain** | 1.009 | 1.002–1.016 |
| **Patient global assessment** | 1.010 | 1.003–1.018 |

95% CI, 95% confidence interval

**Figure 1:** Frequency distribution of SF-6D and EQ-5D utility scores at baseline for 813 patients with early arthritis

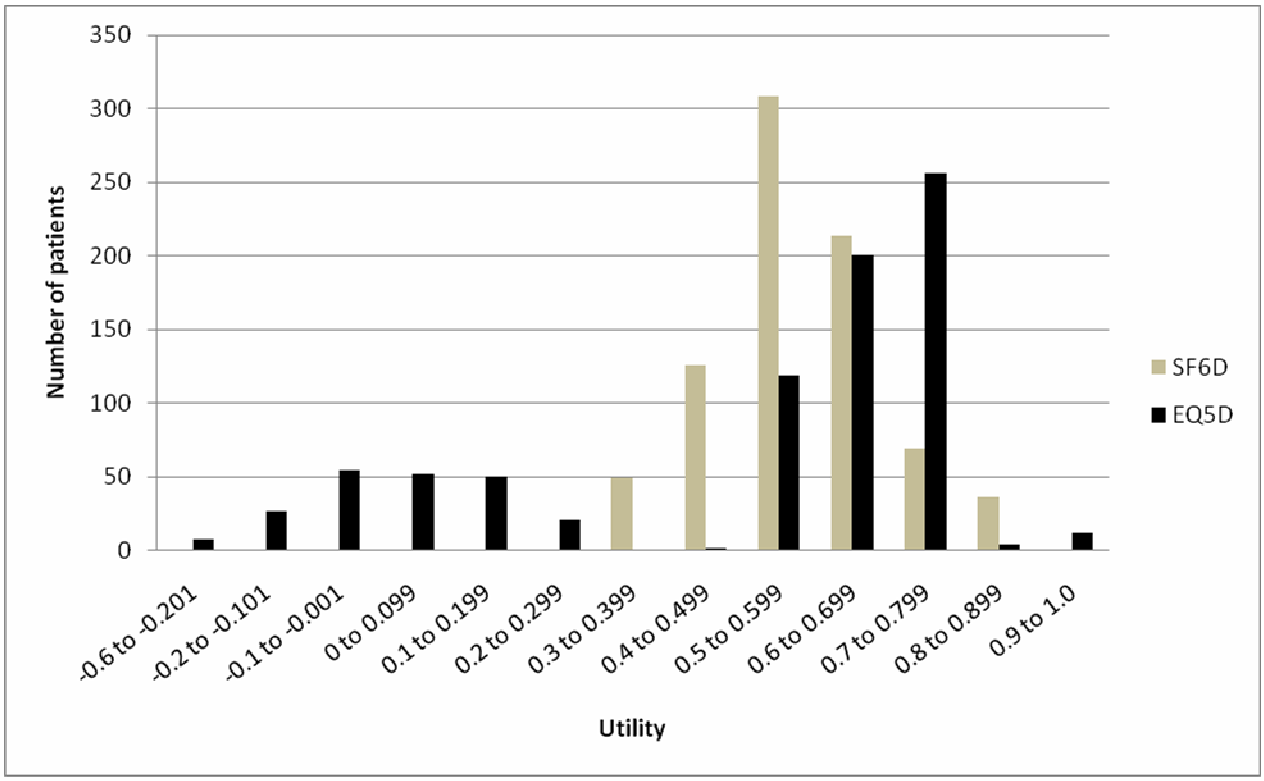
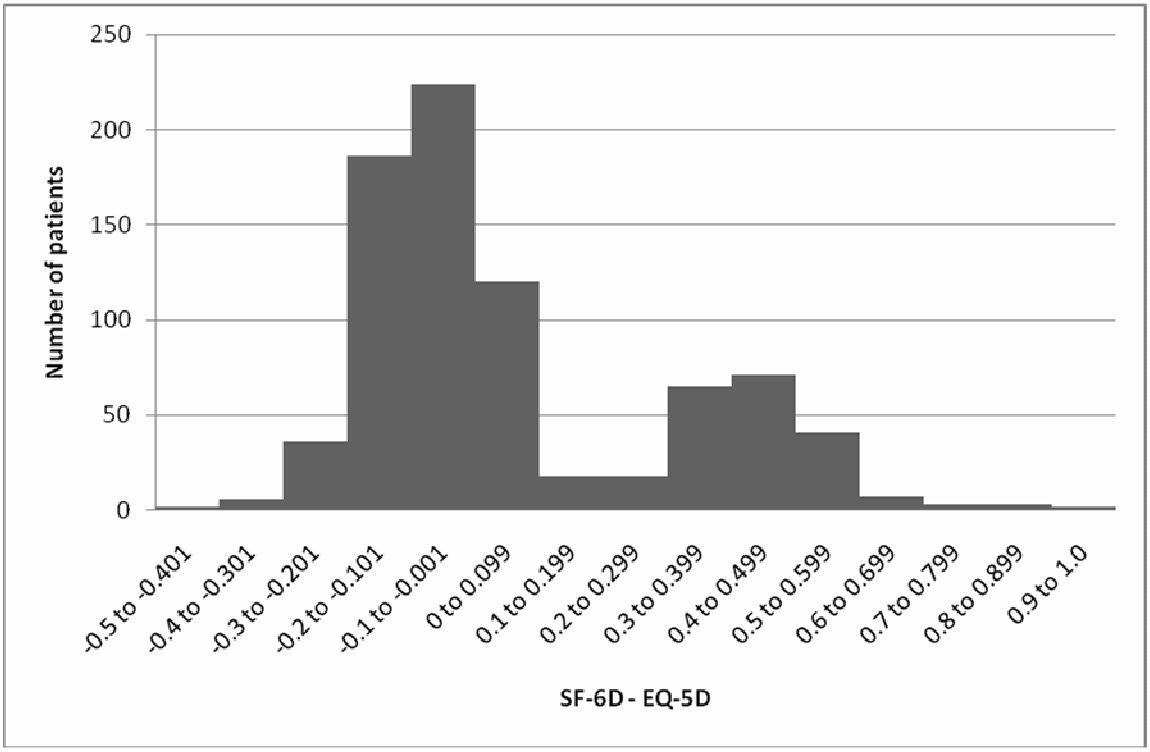


**Figure 2:** Frequency distribution of the difference between SF-6D and EQ-5D scores at baseline for 831 patients with early arthritis



# The determinants of the SF-6D and EQ-5D utility measures vary over time in early arthritis: Results from the ESPOIR cohort

C Gaujoux-Viala[1,2], AC Rat[1,3], K Hosseini[1,3], F Guillemin[1,3], RM Flipo[4], B Fautrel[2]

[1] Lorraine University, Paris Descartes University, EA 4360 Apemac, Nancy, France

[2] Paris 6 – Pierre et Marie Curie University; AP-HP, Department of Rheumatology, Pitié-Salpêtrière Hospital, - GRC-UPMC 08 – EEMOIS, Paris, France

[3] INSERM, CIC-EC CIE6, Nancy, France

[4] Department of Rheumatology, Lille University 2, France

**Key words**: utility, SF-6D, early arthritis

**Institutions to which the work should be attributed**: Lorraine University, Paris Descartes University, EA 4360 Apemac; Paris 6 – Pierre et Marie Curie University; Department of Rheumatology, Pitié-Salpêtrière Hospital, Paris, France

GRC-UPMC 08 – EEMOIS

**Authors:**

Cécile Gaujoux-Viala, MD, Anne-Christine Rat, MD, PhD, Kossar Hosseini, MSc, Francis Guillemin, MD, PhD, René-Marc Flipo, MD, PhD, Bruno Fautrel, MD, PhD

**Corresponding author**: Dr Cécile Gaujoux-Viala,
Université Paris 6 – Pierre & Marie Curie
AP-HP, Groupe hospitalier Pitié-Salpêtrière,
service de Rhumatologie,
83 boulevard de l'Hôpital 75651 Paris cedex 13, France
Phone: +33 1.421.77821 – Fax : +33.1.421.77955
Email: cecilegaujouxviala@yahoo.fr

**Short running footline**: Utility in early arthritis

**Word count**: 3358

**Abstract**. 246 words

Objective: 1) To assess characteristics associated with SF-6D, EQ-5D and with a substantial SF-6D–EQ-5D difference in early arthritis (EA) 2) To check if these associations are stable over 3 years.

Methods: The determinants of SF-6D, EQ-5D and substantial difference (│SF-6D–EQ-5D│>0.074=minimal important difference) at each time-point were assessed by multivariate linear regressions in 813 patients with EA over 3 years.

Results: At baseline, high HAQ score, pain and low AIMS2-SF-mental status were associated with low EQ-5D score explaining 41.4, 0.8 and 7.7% of the EQ-5D variance respectively. At 6 months, 2 and 3 years, results were quite similar. At12 months visit, although the determinants of EQ-5D were similar, EQ-5D was essentially determined by mental status.
At baseline, high HAQ score, pain, fatigue, patient global assessment and low AIMS2-SF-mental status were associated with low SF-6D score. HAQ and AIMS2-SF-mental scores explained 49.3 and 14.5% of the SF-6D variance respectively. Since 6 months visit, the determinants of SF-6D were globally the same. However the part of the variance they explained dramatically changed. HAQ and AIMS2-SF-mental scores explained 7 to 13% and 55 to 57% of the variance respectively. Substantial SF-6D–EQ-5D difference was essentially determined by function, except by pain at 1 year visit.

Conclusion: The major impact of functional ability and mental component, and the variability of the utility determinants over time have consequences for the utility evaluation in EA patients. Evaluation of treatment cost-utility should not be based on utility data transformed from HAQ.

**Introduction.**

The explosion of drug development for rheumatoid arthritis (RA) and the revolution of early aggressive therapy for the disease have fuelled the search for better approaches to establish cost-effectiveness in early arthritis (EA). The utility concept was developed by health economists: utility is the preference (rated in the presence of choice) for a health state relative to perfect health (scored 1) and death (scored 0): an individual considers a transition from a defined health state to some alternative health state which involves a sacrifice of something they value, for example life expectancy, or a risk of death. The greater the sacrifice or risk accepted to make the transition, the lower the valuation of the defined health state (1). Utility may be converted into quality-adjusted life-years (QALY) by multiplying the time spent in a health state by its utility. Cost-utility analysis (or cost per QALY gained) is used for economic evaluation of clinical interventions and health programs (2). It has become increasingly important as health care providers aim to select the treatments and interventions which maximize health gain from their scarce resources. While methods ascertained by face-to-face interview such as the standard gamble (SG) and the time trade-off (TTO) may be used to measure utility directly (1,3,4), they are less suitable for clinical research and less widely used for feasibility reasons. Therefore, indirect utility assessment techniques such as EQ-5D (5) and SF-6D (6) have been developed. These multi-attribute utility measures are estimated from generic Health Related Quality Of Life (HRQoL) instruments to which one applies the weight assigned by the general population for each particular health state (7).

Although the theoretical concept of utility implies that one specific health state has one utility score, regardless of how it is measured, EQ-5D and SF-6D may generate disparate utilities in the same patients as shown in RA (8,9) and in EA (10). Furthermore in studies where no preference based measures are collected, indirect methods have been developed to estimate utilities from clinical instruments (11,12). Mapping is one solution that is gaining popularity as it enables health state utility values to be predicted when no preference-based measure has been

included in the study. This approach involves estimating the relationship between a non-preference-based measure and a generic preference based measure using statistical association (also known as 'cross-walking') (13).

Here we aimed to assess, in a large group of patients with EA 1) the socio-demographic and clinical characteristics associated with SF-6D and EQ-5D scores at different time-points over 3 years 2) which of these variables are associated with substantial difference between the scores 3) to check if these associations are stable over 3 years.

## Material and Methods.

### *Patients*

Between December 2002 and March 2005, 813 patients with EA from 14 French regional centers were included in the ESPOIR cohort (14). Inclusion criteria were age 18 to 70 years, more than 2 swollen joints for >6 weeks and <6 months, suspected or confirmed diagnosis of RA, and not taking disease-modifying antirheumatic drugs or steroids (except if < 2 weeks). Patients were excluded if the referring physician judged they had other clearly defined inflammatory rheumatic diseases. Patients were followed every 6 months during the first 2 years then every year. At baseline and at each visit, data for a set of clinical and biological variables were recorded, by the Disease Activity Score for 28 Joints (DAS28), a composite index of disease activity (range: 0.49 -9.07 : with 9.07 high activity) (15), a functional ability questionnaire, the Health Assessment Questionnaire (HAQ) (range: 0-3, with 3 the worst incapacity) (16), and pain at rest on a 100-mm visual analogue scale (range: 0-100, with 100 the worst pain). Fatigue severity over the previous 48 hours was assessed by a 100-mm visual analog scale, (range: 0-100, with 100 the worst fatigue). Baseline C-reactive protein level, IgM and IgA rheumatoid factor (ELISA, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive > 50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

Furthermore, patients completed other self-administered patient-reported outcome measures, including health-related quality of life questionnaires, the Arthritis Impact Measurement Scale 2–Short Form (AIMS2-SF) (17), the EQ-5D (18) and the SF-36 (19). The protocol of the ESPOIR Cohort study was approved by the ethics committee of Montpellier, France. All patients gave their signed informed consent before inclusion.

*Utility measurement*

*EQ-5D*

The EQ-5D comprises of 5 items or dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). For each dimension there are three answering categories: no problem (1), some problems (2), or severe problems (3), resulting in 243 ($3^5$) health states. The EQ-5D can be reported in terms of a 5-digit profile indicating the level on each dimension, or in terms of a preference-based single index number. The latter is obtained by applying algorithm that link the 5-digit health state description to mean values obtained from members of the general population in the UK (n=3395) using the TTO method. This algorithm yields a range from -0.59 to +1, with 1.00 representing full health, 0 death and negative value, a health state labelled “worse than death” (18).

*SF-6D*

The SF-6D is derived from the widely used and validated SF-36 generic health survey. Eleven questions of the SF-36 are used to create the 6 health domains of the SF-6D: physical functioning, role limitation, social functioning, pain, mental health, and vitality, with 4–6 levels for each domain. A SF-6D health state is defined by 1 level from each domain, which results in a total of 18,000 possible health states. Health state valuations were derived using a representative sample of the UK population (n=836) using the SG technique. Utility scores generated by the SF-6D range from 0.29 to 1.00, with 1.00 representing full health and 0.29

representing the worst possible health state defined by the SF-6D (i.e., all domains being at the worst level) (6).

### *Statistical analysis*

#### *Global utility scores*

Descriptive statistics (mean [standard deviation], median, minimum, maximum) and distributions of the EQ-5D and the SF-6D utility scores at baseline were described. Variables potentially associated with the EQ-5D and with the SF-6D were analysed by bivariate analysis. Those significant at 0.10 level or less were selected and included in a multivariate linear regression (stepwise selection) to assess which baseline variables are associated with the EQ-5D and SF-6D scores in EA. Partial R-squares were used to assess the relative importance of variables to the variation in EQ-5D and SF-6D scores. All analyses were repeated at 6, 12, 24, and 36 months to check the stability of the models.

#### *Difference between the SF-6D and the EQ-5D*

Descriptive statistics (mean [standard deviation], median, minimum, maximum) and distributions of the difference between the 2 utility measures (i.e. SF-6D – EQ-5D) at baseline were computed.

To determine which variables are associated with a substantial difference between the 2 instruments, we chose a cutoff of 0.074 between the scores on the basis of the high estimate of the minimal important difference (MID) published for the EQ-5D and SF-6D (20). Walters et al. compared the MID for the SF-6D and EQ-5D from 8 longitudinal studies that used both instruments with 11 patient groups. The mean MID for the SF-6D was 0.041 and for the EQ-5D, 0.074 (20). Thus, we considered that patients with $|\text{SF-6D} - \text{EQ-5D}| > 0.074$ showed a substantial difference between the 2 scores. In these patients, variables potentially associated

with the difference between the 2 utility scores at baseline were analysed by bivariate analysis. Those significant at 0.10 level or less were selected and included in a multivariate linear regression (stepwise selection) to assess which baseline variables were associated with substantial difference between the 2 scores in EA. Partial R-square was used to explain the relative importance of individual variables to the substantial difference between the SF-6D and EQ-5D scores.

Due to heterogeneity in potential MIDs, we performed a sensitivity analysis with a cut-off for a substantial difference of 0.129 on the basis of the higher estimate of the MID published for the EQ-5D and SF-6D in early RA (20). We also use a cut-off of 0.06: the MID for SF-6D was estimate around 0.03 in several studies (20,21) and the EQ-5D scale having approximately twice the range of the SF-6D scale so the estimate of the MID for the EQ-5D is almost double that of the estimate of the MID for the SF-6D (20).

All analyses were repeated at 6, 12, 24, and 36 months to check the stability of the models.

All analyses involved use of SAS v9.2 (SAS Inst. Inc., Cary, NC, USA). A $p<0.05$ was considered statistically significant.

## Results

*Characteristics of the population*

Table 1 shows the demographic and clinical characteristics of the 813 patients in the ESPOIR cohort at inclusion. In total, 641 (78.8%) patients fulfilled the new ACR-EULAR criteria for RA. At 3 years, 618 patients were still being followed up: their demographic and clinical characteristics are detailed in the table 1. Baseline results were similar when we considered only the 618 patients still present at 3 years.

*Global utility scores*

*EQ-5D*

The distribution of EQ-5D scores was bimodal. At baseline, the mean score was 0.52 ± 0.31, median 0.66 [range - 0.59 to 1]. Few missing values were observed: 0.6%.

By multivariate linear regression, high HAQ (0-3: the worst), high pain and low mental status (AIMS2-SF) (0-100: the best) were significant determinants of a low EQ-5D at baseline. HAQ and AIMS2-SF-mental scores explained 41.4 and 7.9% of the variance respectively (Table 2).

At 6 months, 2 and 3 years, high HAQ, high fatigue, high patient global evaluation, and low mental status (AIMS2-SF) were determinants of a low EQ-5D. HAQ and AIMS2-SF-mental scores explained 37.2 to 45 and 7.9 to 11.7% of the variance respectively (Table 2).

At 1 year, the determinants of EQ-5D were almost the same. However the part of the variance they explained dramatically changed (Figure 1): HAQ and AIMS2-SF-mental scores explained 11.4% and 41.7 of the variance of the models respectively (Table 2).

*SF-6D*

The distribution of SF-6D scores was near normal. At baseline, the mean utility score for the SF-6D was 0.58 ± 0.11, median 0.58 [range 0.30 to 0.92]. Few missing values were observed: 1.2%.

By multivariate linear regression, high HAQ score, pain at rest, patient global evaluation and fatigue, and low mental status and educational level were significantly associated with low SF-6D score. HAQ and AIMS2-SF-mental scores explained 49.3 and 14.5% of the variance respectively (table 3). At 6 months visit and further time-points, the determinants of SF-6D were almost the same. Compared to baseline results, the part of the variance they explained dramatically changed (Figure 1). HAQ and AIMS2-SF-mental scores explained 7 to 13% and 55 to 57% of the variance of the models respectively (Table 3).

*Difference between the SF-6D and the EQ-5D*

The distribution of the difference between the 2 utility scores was bimodal at baseline and a bit skewed at 3 years (Figure 2).

The mean difference in utility scores for the 2 measures was mean 0.064 [range -0.421;0.941]) at baseline and – 0.018 ± 0.184 [range -0.368;0.651] at 3 years. At baseline, most patients (66.5%) showed a substantial difference between the 2 scores (MID > 0.074). For these patients, high HAQ score, pain and activity, and low mental status explained the difference between the 2 scores. HAQ and AIMS2-SF mental health scores explained 43.3% and 2% of the variance, respectively (Table 4).

At 3 years, 68% of patients showed a substantial difference between the 2 scores: it was explained by high HAQ score and low mental status. HAQ and AIMS2-SF-mental scores explained 21.8 and 2.2% of the variance respectively.

Substantial difference between SF-6D and EQ-5D was essentially determined by function except at 1 year visit where it was essentially determined by pain (Table 4). However the different models only explained 19 to 47 % of the variance (Table 4).

Due to heterogeneity in potential MIDs, we performed a sensitivity analysis with cut-off for a substantial difference of 0.129 and 0.06. Results were similar.

## Discussion

This study is the first to examine in a real life setting the determinants of the EQ-5D, the SF-6D utility measures and their difference in EA patients over a long period of time, 3 years. At baseline, SF-6D was essentially determined by function whereas after 6 months, SF-6D was essentially determined by mental status. At each time-points, EQ-5D was essentially determined by function except at 1 year, it was essentially determined by mental status. Substantial difference between SF-6D and EQ-5D was essentially determined by function except at 1 year visit it was essentially determined by pain. The major impact of functional

ability and mental component, and the variability of the utility determinants over time have consequences for the utility evaluation in EA patients.

Except at 1 year, EQ-5D was essentially determined by function. This result may be explained by the content of EQ-5D. Of the five dimensions, four (mobility, self-care, usual activity and pain/discomfort) are likely to be particularly affected in EA patients with decreased functional abilities.

At baseline, SF-6D was essentially determined by function whereas after 6 months, SF-6D was essentially determined by mental status. It seems that since the incapacity improves, the mental aspect becomes dominating. In a previous work, we have shown that the correlations between scores of the mental component of the SF-36, and the SF-6D or the EQ-5D increased at 6 months and remained stable thereafter (10). The increase of the correlations could be explained by the importance, the emergence of the mental health component when the incapacity decreases and how patients deal with and are able to cope with a recent diagnosis of a chronic disease.

Substantial difference between the 2 utility measures in EA patients was essentially determined by high functional disability except at 1 year, it was essentially determined by pain. However the models only explained 19 to 47% of the variance.

The EQ-5D allows for negative utility values, which theoretically correspond to health states worse than death, not available with other indirect utility instruments (like SF-6D…). In a precedent study, we analyzed patients with EQ-5D < 0 in the ESPOIR cohort (22). At baseline, 90 (11%) patients showed EQ-5D < 0. This proportion was decreased at 6 months but remained stable over time, at 3.3%. Pain/discomfort was the key EQ-5D domain associated with a state worse than death and was extreme for most patients with EQ-5D < 0, especially at baseline. On multivariate analysis, the HAQ score and SF-36 mental component score were independently and consistently associated with a state worse than death over 2 years (22).

Functional disability is one of the main drivers of the difference between the EQ-5D and the SF-6D. A study comparing EQ-5D and SF-6D in seven diseases, showed high mean differences between the two instruments in osteoarthritis than in diseases focusing on pain and discomfort such as irritable bowel syndrome (23). Differences between the utility scores may also be confounded by the different operational definitions of the domains and functional levels within each domain used by the 2 instruments.

As not all studies choose to measure outcomes in terms of utility, an increasing amount of research has now been conducted on mapping, where scores from a condition-specific (non preference-based) measure are 'converted' into a utility (preference-based) score using a pre-defined formulae. Mapping thereby presents the possibility of estimating the cost-utility ratio (i.e. the incremental cost per QALY) of interventions that have only been evaluated using a condition-specific measure for effectiveness. A number of mapping models have now been developed (12) and the use of mapping has been considered by the UK National Institute of Health and Clinical Excellence (NICE). The HAQ was invariably recorded in clinical trials in RA. Subsequently, linear transformations between the HAQ and utility have previously been used to estimate the utility scores (11), and in turn cost-utility, of a number of health care interventions (24,25). Amjadi et al, assessed the validity of SF-6D scores obtained by the methods described by Bansback et al (11) and found that SF-6D scores resulting from HAQ transformation were valid in terms of construct validity, discriminant capacity and responsiveness to change in an early RA cohort (26). However, in observational studies of RA patients, Scott et al concluded that EQ-5D and HAQ were moderately inversely correlated (27). The distribution was Gaussian for the HAQ and bimodal for the EQ-5D. In the treatment change study, changes in HAQ and EQ-5D were unrelated (27). In a previous study, we found good and stable correlations between EQ-5D or SF-6D and HAQ in the ESPOIR cohort over 2 years (Spearman correlations coefficients = -0.70) (10). Despite these stable correlations over time, we have demonstrated in this study the variability of the utility determinants over time:

SF-6D was essentially determined by function at baseline whereas after 6 months, it was essentially determined by mental status. At each time-points, EQ-5D was essentially determined by function except at 1 year visit, it was also essentially determined by mental status. So good construct validity is not sufficient to support the use of mapping. Moreover Harrison et al compared predicted and observed EQ-5D and SF-6D scores using data from 3 cohorts of patients with EA, established stable RA and RA being treated with anti TNF. They found that predicted utility scores provided overestimated values at baseline value but underestimated values of change (28). Brazier et al conducted recently a systematic review of current practice in mapping between non-preference based measures and generic preference based measures (12). They concluded that the performance of the mapping functions in terms of goodness-of-fit and prediction was variable and generalization was not possible across instruments. Further work is required to test the accuracy of these functions in more relevant contexts and over a larger range of instruments. However, the use of mapping to derive preference-based generic indices from condition-specific measures raises a more fundamental concern. Mapping assumes that there is a conceptual overlap between what is being measured by the two measures, and further that the preference-based target measure covers all important aspects of health of the non preference- based start measure. In other words, the strength of the mapping function is related to the degree of overlap in content between the instruments being mapped. Where there are important dimensions of one instrument not covered by the other, this may undermine the model and could be an important weakness (12).

Our study has some limitations. The EQ-5D modelisation was hampered because the EQ-5D scores showed a high level of skewness as compared with the normal distribution of the SF-6D scores. The scoring algorithms used for both instruments were developed from data for a general population in the United Kingdom because no such algorithm was available in France for the SF-6D at the time of the study. Such an algorithm for the EQ5D has been developed in

France (29) but the use of algorithms from the same population for both the EQ-5D and SF-6D might result in a more valid comparison.

One of the strengths of our study is its investigation of a wide range of EA states/severity found in routine practice. The observational ESPOIR cohort aimed to include all patients with EA regardless of disease level, age and sex, and our study implied the performance of the instruments in a real-life setting. The study also included a large number of patients with longitudinal assessment and 5 repeated measures over 3 years.

This study aimed to focus on developing a fuller understanding of the determinants of the SF-6D, the EQ-5D and of the difference between the SF-6D and the EQ-5D in EA over time. Function and mental state were the main determinants of SF-6D and EQ-5D utility measures. High functional disability and low mental status were the key element leading to substantial difference between the 2 utility measures in EA patients. The variability of the determinants of SF-6D and EQ-5D utility measure over 3 years in addition of the bimodal distribution of the EQ-5D raise concerns about mapping to estimate utilities from clinical instruments. Evaluation of treatment cost utility should not be based on SF-6D or EQ-5D data transformed from HAQ. Future clinical studies should include at least one preference based measure of utility.

**Acknowledgement**

"An unrestricted grant from Merck Sharp and Dohme (MSD) was allocated for the first 5 years of the ESPOIR cohort study. Two additional grants from INSERM were obtained to support part of the biological database. The French Society of Rheumatology, Abbott, and Wyeth-Pfizer also supported the ESPOIR cohort study.

We also wish to thank Nathalie Rincheval who did expert monitoring and data management and all the investigators who recruited and followed the patients (F. BERENBAUM, Paris-Saint Antoine, MC. BOISSIER, Paris-Bobigny, A .CANTAGREL, Toulouse, B. COMBE, Montpellier, M. DOUGADOS, Paris-Cochin, P FARDELONNE, Amiens, B. FAUTREL, P BOURGEOIS, Paris-La Pitié, RM. FLIPO, Lille, Ph. GOUPILLE, Tours, F. LIOTE, Paris-Lariboisière, X. LE LOET Rouen, X MARIETTE, Paris Bicêtre, O MEYER, Paris Bichat, SARAUX, Brest, Th SCHAEVERBEKE, Bordeaux, J. SIBILIA, Strasbourg)."

**References**


1. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986 mars;5(1):1-30.

2. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. N. Engl. J. Med. 1977 mars 31;296(13):716-21.

3. Bakker C, van der Linden S. Health related utility measurement: an introduction. J. Rheumatol. 1995 juin;22(6):1197-9.

4. Drummond MF, Sculpher MJ, Torrance GW. Methods for economic evaluation of health care programmes. Oxford: Oxford Medical Publications; 2005.

5. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy. 1990 déc;16(3):199-208.

6. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271-92.

7. Khanna D, Tsevat J. Health-related quality of life--an introduction. Am J Manag Care. 2007 déc;13 Suppl 9:S218-223.

8. Lillegraven S, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. Ann. Rheum. Dis. 2010 oct;69(10):1762-7.

9. Marra CA, Marion SA, Guh DP, Najafzadeh M, Wolfe F, Esdaile JM, et al. Not all « quality-adjusted life years » are equal. J Clin Epidemiol. 2007 juin;60(6):616-24.

10. Gaujoux-Viala C, Rat A-C, Guillemin F, Flipo R-M, Fardellone P, Bourgeois P, et al. Comparison of the EQ-5D and the SF-6D utility measures in 813 patients with early arthritis: results from the ESPOIR cohort. J. Rheumatol. 2011 août;38(8):1576-84.

11. Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T, et al. Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. Arthritis Rheum. 2007 août 15;57(6):963-71.

12. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ. 2010 avr;11(2):215-25.

13. Fryback DG, Dasbach EJ, Klein R, Klein BE, Dorn N, Peterson K, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making. 1993 juin;13(2):89-102.

14. Combe B, Benessiano J, Berenbaum F, Cantagrel A, Daures JP, Dougados M, et al. The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. Joint Bone Spine. 2007;74(5):440-5.

15. Prevoo ML, van ’t Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum. 1995;38(1):44-8.

16. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23(2):137-45.

17. Guillemin F, Coste J, Pouchot J, Ghézail M, Bregeon C, Sany J. The AIMS2-SF: a short form of the Arthritis Impact Measurement Scales 2. French Quality of Life in Rheumatology Group. Arthritis Rheum. 1997 juill;40(7):1267-74.

18. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997 nov;35(11):1095-108.

19. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992 juin;30(6):473-83.

20. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005 août;14(6):1523-32.

21. Gandhi PK, Ried LD, Bibbey A, Huang I-C. SF-6D utility index as measure of minimally important difference in health status change. J Am Pharm Assoc (2003). 2012 févr;52(1):34-42.

22. Gaujoux-Viala C, Fautrel B, Guillemin F, Flipo R-M, Bourgeois P, Rat A-C. Who are the patients with early arthritis with “worse than death” scores on the EQ-5D? Results from the ESPOIR cohort. Rheumatology 2012 [Epub ahead of print]

23. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004 sept;13(9):873-84.

24. Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. Health Technol Assess. 2004 mars;8(11):iii, 1-91.

25. Brennan A, Bansback N, Reynolds A, Conway P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. Rheumatology (Oxford). 2004 janv;43(1):62-72.

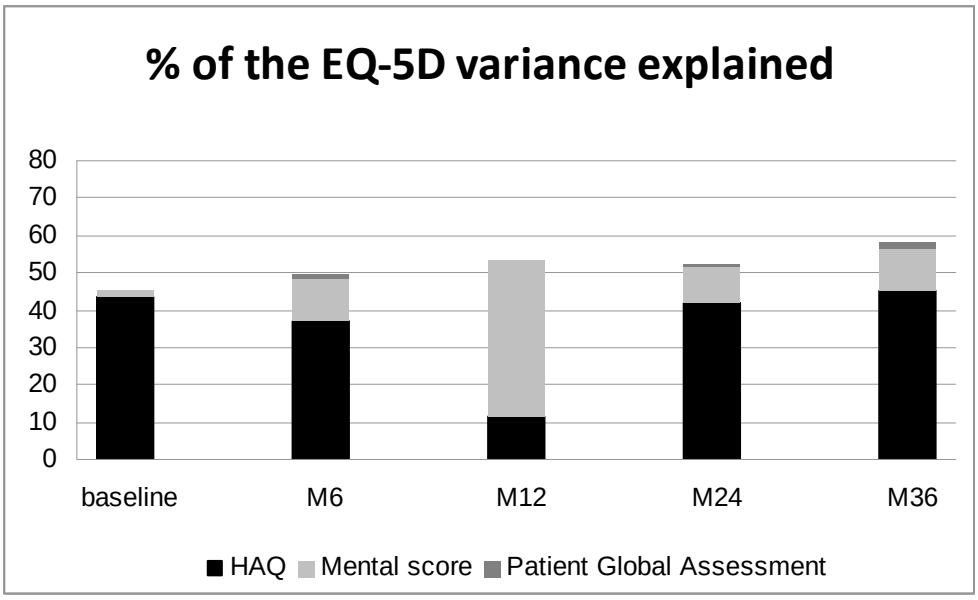
26. Amjadi SS, Maranian PM, Paulus HE, Kaplan RM, Ranganath VK, Furst DE, et al. Validating and assessing the sensitivity of the Health Assessment Questionnaire-Disability Index-derived Short Form-6D in patients with early aggressive rheumatoid arthritis. J. Rheumatol. 2009 juin;36(6):1150-7.

27. Scott DL, Khoshaba B, Choy EH, Kingsley GH. Limited correlation between the Health Assessment Questionnaire (HAQ) and EuroQol in rheumatoid arthritis: questionable validity of deriving quality adjusted life years from HAQ. Ann. Rheum. Dis. 2007 nov;66(11):1534-7.

28. Harrison MJ, Lunt M, Verstappen SMM, Watson KD, Bansback NJ, Symmons DPM. Exploring the validity of estimating EQ-5D and SF-6D utility values from the health assessment questionnaire in patients with inflammatory arthritis. Health Qual Life Outcomes. 2010;8:21.

29. Chevalier J, de Pouvourville G. Valuing EQ-5D using Time Trade-Off in France. Eur J Health Econ. 2011 Sep 21. [Epub ahead of print]

**Figure 1:** Evolution of the percentage of the EQ-5D and SF-6D variance explained in EA patients over 3 years



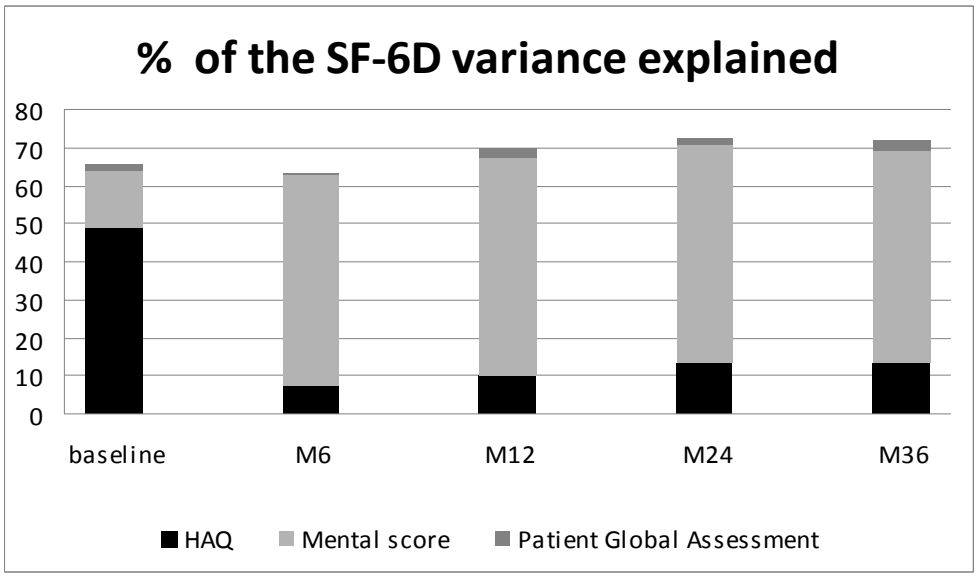



**Figure 2:** Frequency distribution of the difference between SF-6D and EQ-5D at baseline and 3 years

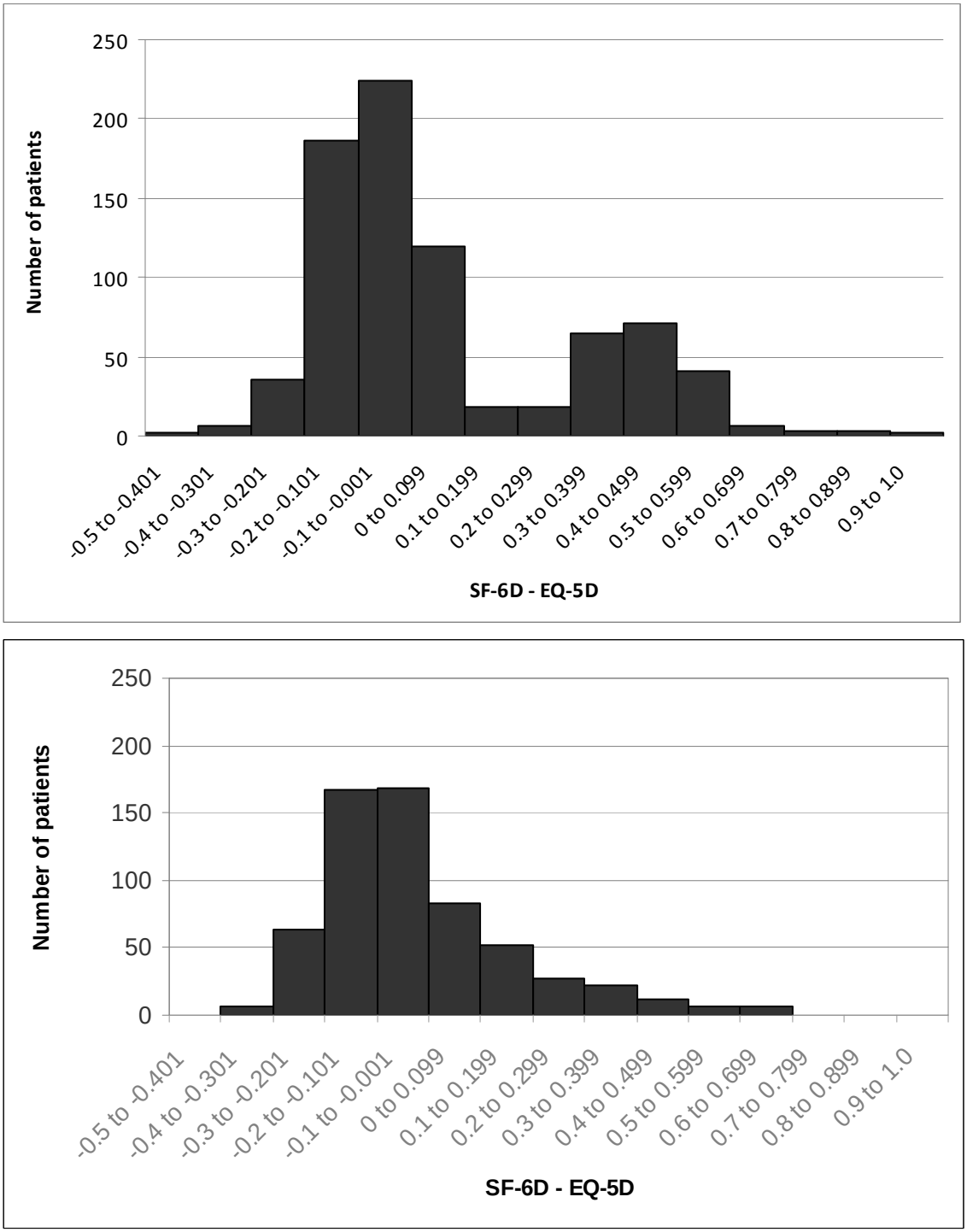


**Table 1:** Characteristics of patients included in ESPOIR cohort at baseline and 3 years

| Characteristic | **All patients at baseline (n=813)** Mean ± SD | **All patients at 6 months (n=757)** Mean ± SD | **All patients at 3 years (n=618)** Mean ± SD |
|---|---|---|---|
| Age (years) | 48.1 ± 12.6 | 48.8 ± 12.5 | 51.7 ± 12.0 |
| Female sex, N (%) | 624 (76.7) | 578 (76) | 475 (76.9) |
| >12 years of education, N (%) | 255 (31.4) | 244 (32.2) | 200 (32.3) |
| Married or living together, N (%) | 594 (73.1) | 573 (75.7) | 464(75.1) |
| DAS28 | 5.11 ± 1.31 | 3.39 ± 1.40 | 2.87 ± 1.36 |
| HAQ (0-3: with 3 high disability) | 0.979 ± 0.684 | 0.548 ± 0.566 | 0.499 ± 0.580 |
| AIMS2-SF-mental score (0-100, with 100 the best mental status) | 58.3 ± 19.9 | 66.2 ± 20.4 | 69.3 ± 20.2 |
| Patient Global Assessment (0-100, with 100 the worst health) | 59.8 ± 25.6 | 34.7 ± 26.8 | 29.0 ± 26.2 |
| Pain at rest (0-100, with 100 high pain) | 37.2 ± 27.6 | 23.3 ± 24.0 | 19.6 ± 23.0 |
| Fatigue (0-100, with 100 the worst fatigue) | 47.8 ± 28.2 | 39.0 ± 28.4 | 33.6 ± 28.7 |
| Erythrocyte sedimentation rate | 29.4 ± 24.6 | 16.2 ± 14 | 14.9 ± 13.3 |
| C-reactive protein level * | 20.3 ± 32.4 | 8.8 ± 19.9 | 6.8 ± 11.3 |
| Rheumatoid factor*, N (%) | 372 (45.8) | 347 (45.8) | 300 (48.5) |
| Anti-CCP2 antibodies*, N (%) | 315 (38.8) | 296 (39) | 268 (43.4) |
| V H Sharp score $ | 5.7 ± 7.70 | NA | 13.9 ± 14.7 |
| SF-6D score | 0.58 ± 0.11 | 0.67 ± 0.12 | 0.69 ± 0.13 |
| EQ-5D score | 0.52 ± 0.31 | 0.66 ± 0.23 | 0.70 ± 0.24 |

NA: not available DAS28: disease activity score in 28 joints; HAQ: Health Assessment Questionnaire; AIMS2-SF: Arthritis Impact Measurement Scale 2–Short Form

*Baseline C-reactive protein level (normal<10 mg/l); IgM and IgA rheumatoid factor (Elisa, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive > 50 U/ml) were detected in all patients using the same technique in a central lab (Paris-Bichat).

$ Van der Heijde modified Sharp score.

**Table 2:** Determinants of the EQ-5D utility scores in EA patients at baseline and 3 years by multivariate linear regression

| Model | Variables | Beta | p | Partial $R^2$ |
|---|---|---|---|---|
| **EQ-5D at baseline** | HAQ | -0.20 | <0.0001 | 41.4% |
| | (0-3 with 3 the worst disability) | | | |
| $R^2$=0.51 | AIMS2-SF-mental score | 0.005 | <0.0001 | 7.9% |
| | (0-100 with 100 the best mental status) | | | |
| | Pain at rest | -0.0009 | 0.002 | 0.84% |
| | (0-100 with 100 high pain) | | | |
| **EQ-5D at M6** | HAQ | -0.13 | <0.0001 | 37.2% |
| | AIMS2-SF-mental score | 0.004 | <0.0001 | 11.5% |
| $R^2$=0.52 | Pain at rest | -0.0009 | 0.0046 | 0.18% |
| | Patient global assessment | -0.0009 | 0.0067 | 0.6% |
| | (0-100 with 100 the worst health) | | | |
| | Educational level $ | 0.026 | 0.022 | 0.28% |
| | Married or living together | 0.030 | 0.046 | 0.33% |
| **EQ-5D at M12** | HAQ | -0.13 | <0.0001 | 11.4% |
| | AIMS2-SF-mental score | 0.005 | <0.0001 | 41.7% |
| $R^2$=0.56 | Pain at rest | -0.002 | <0.0001 | 3.25% |
| **EQ-5D at M24** | HAQ | -0.14 | <0.0001 | 42.1% |
| | AIMS2-SF-mental score | 0.004 | <0.0001 | 9.4% |
| $R^2$=0.55 | Pain at rest | -0.001 | <0.0001 | 2.26% |
| | Patient global assessment | -0.0008 | 0.012 | 0.5% |
| | Female | -0.034 | 0.043 | 0.32% |
| **EQ-5D at M36** | HAQ | -0.14 | <0.0001 | 45% |
| | AIMS2-SF-mental score | 0.004 | <0.0001 | 11.7% |
| $R^2$=0.59 | Patient global assessment | -0.001 | <0.0001 | 1.4% |
| | Fatigue | -0.0008 | 0.0083 | 0.5% |
| | (0-100 with 100 the worst) | | | |

$ >12 years of education

**Table 3:** Determinants of the SF-6D in EA patients over 3 years by multivariate linear regression

| Model | Variables | Beta | p | Partial $R^2$ |
|---|---|---|---|---|
| **SF-6D at baseline** $R^2$=0.66 | HAQ (0-3 with 3 the worst disability) | -0.072 | <0.0001 | 49.3% |
| | AIMS2-SF-mental score (0-100 with 100 the best mental status) | 0.0023 | <0.0001 | 14.5% |
| | Patient global assessment (0-100 with 100 the worst health) | -0.0005 | <0.0001 | 1.8% |
| | Pain at rest (0-100 with 100 high pain) | -0.0002 | 0.0005 | 0.5% |
| | Fatigue (0-100 with 100 the worst fatigue) | -0.0002 | 0.0186 | 0.2% |
| | Educational level $ | 0.011 | 0.035 | 0.2% |
| **SF-6D at M6** $R^2$=0.66 | HAQ | -0.049 | <0.0001 | 7.4% |
| | AIMS2-SF-mental score | 0.0030 | <0.0001 | 55.3% |
| | Patient global assessment | -0.0006 | <0.0001 | 0.9% |
| | Age | -0.0006 | 0.0008 | 0.5% |
| | Fatigue | -0.0006 | <0.0001 | 2.1% |
| **SF-6D at M12** $R^2$=0.72 | HAQ | -0.055 | <0.0001 | 10.3% |
| | AIMS2-SF-mental score | 0.003 | <0.0001 | 57% |
| | Patient global assessment | -0.0006 | <0.0001 | 2.4% |
| | Age | -0.001 | <0.0001 | 0.91% |
| | Fatigue | -0.0006 | <0.0001 | 0.91% |
| | Pain at rest | -0.0003 | 0.025 | 0.2 % |
| **SF-6D at M24** $R^2$=0.75 | HAQ | -0.073 | <0.0001 | 13.4% |
| | AIMS2-SF-mental score | 0.0029 | <0.0001 | 57.3% |
| | Patient global assessment | -0.00038 | <0.0001 | 1.5% |
| | Age | -0.00074 | 0.0005 | 0.5% |
| | Fatigue | -0.00052 | 0.0002 | 0.6% |
| | DAS28 | -0.0064 | 0.0208 | 0.2% |
| **SF-6D at M36** $R^2$=0.73 | HAQ | -0.069 | <0.0001 | 13.2% |
| | AIMS2-SF-mental score | 0.0028 | <0.0001 | 56.4% |
| | Patient global assessment | -0.00075 | <0.0001 | 2.3% |
| | Age | -0.001 | <0.0001 | 0.1% |
| | Fatigue | -0.00058 | 0.0001 | 0.7% |

$ >12 years of education

**Table 4:** Determinants of difference between the 2 utility measures in EA patients with substantial difference at different time-points over 3 years by multivariate linear regression

| **Model**<br>Number of patients with substantial difference<br>*Mean difference ± SD in all patients* | Variables | Beta | p | Partial $R^2$ |
|---|---|---|---|---|
| **\|SF-6D–EQ5D\| at baseline**<br>$R^2$=0.47<br>N= 541<br>*0.063 ± 0.247* | HAQ (0-3 with 3 the worst disability)<br>AIMS2-SF-mental score (0-100 with 100 the best mental status)<br>Log CRP level | 0.208<br>-0.0023<br>0.036 | <0.0001<br><0.0001<br>0.0005 | 43.3%<br>2%<br>1.4% |
| **\|SF-6D–EQ5D\| at M6**<br>$R^2$=0.19<br>N= 424<br>*0.0094 ± 0.180* | HAQ<br>Patient global assessment<br>AIMS2-SF-mental score | 0.0425<br>0.0009<br>-0.0008 | <0.0001<br>0.0002<br>0.024 | 14.3%<br>3.3%<br>1.2% |
| **\|SF-6D–EQ5D\| at M12**<br>$R^2$=0.23<br>N= 424<br>*-0.0016 ± 0.183* | Pain at rest (0-100 with 100 high pain)<br>AIMS2-SF-mental score<br>HAQ | 0.001<br>-0.0012<br>0.041 | <0.0001<br><0.0001<br>0.0019 | 16%<br>4.9%<br>2% |
| **\|SF-6D–EQ5D\| at M24**<br>$R^2$= 0.26<br>N= 377<br>*-0.0021 ± 0.180* | HAQ<br>AIMS2-SF-mental score<br>Pain at rest | 0.195<br>-0.0044<br>0.0036 | <0.0001<br>0.0011<br>0.0118 | 17%<br>2.8%<br>1.6% |
| **\|SF-6D–EQ5D\| at M36**<br>$R^2$= 0.24<br>N= 420<br>*-0.018 ± 0.184* | HAQ<br>AIMS2-SF-mental score | 0.0625<br>-0.001 | <0.0001<br>0.0012 | 21.8%<br>2.23% |