



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>

28 MARS 2001

01302
DOUBLE
158483

UNIVERSITE HENRI POINCARE, NANCY I

FACULTE DE MEDECINE DE NANCY

THESE DE DOCTORAT DE L'UNIVERSITE HENRI POINCARE, NANCY I

BIOLOGIE, SANTE, ENVIRONNEMENT

Spécialité: Epidémiologie et Santé Publique



**CONTRIBUTION A L'OPTIMISATION D'UN CRITERE DE JUGEMENT :
MESURE DE L'INTERLIGNE ARTICULAIRE DANS LA COXARTHROSE**

THESE

Présentée et soutenue publiquement

le 26 janvier 2001

pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE HENRI POINCARE, NANCY I

Par

AULELEY Guy-Robert

Né le 30/07/1961

A Lambaréné (Gabon)

JURY : Mr le Professeur Serge BRIANÇON- Président du Jury (Nancy)
Mme le Professeur Isabelle CHARY-VALCKENAERE (Nancy)
Mr le Professeur Maxime DOUGADOS (Paris)
Mr le Professeur Francis GUILLEMIN- Directeur de Thèse (Nancy)
Mr le Professeur Bernard MAZIERES- Rapporteur (Toulouse)
Mr le Professeur Philippe RAVAUD- Directeur de Thèse (Paris)

RAPPORTEURS : Mr le Professeur Bernard MAZIERES (Toulouse)
Mr le Professeur Gilles CHATELLIER (Paris)

BIBLIOTHEQUE MEDECINE NANCY 1



D

007 209208 9

**THESE DE DOCTORAT DE L'UNIVERSITE HENRI POINCARÉ, NANCY I
BIOLOGIE, SANTE, ENVIRONNEMENT**

Spécialité: Epidémiologie et Santé Publique

**CONTRIBUTION A L'OPTIMISATION D'UN CRITERE DE JUGEMENT :
MESURE DE L'INTERLIGNE ARTICULAIRE DANS LA COXARTHROSE**

THESE

Présentée et soutenue publiquement

le 26 janvier 2001

pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE HENRI POINCARÉ, NANCY I

Par

AULELEY Guy-Robert

Né le 30/07/1961

A Lambaréné (Gabon)

JURY : Mr le Professeur Serge BRIANÇON- Président du Jury (Nancy)
Mme le Professeur Isabelle CHARY-VALCKENAERE (Nancy)
Mr le Professeur Maxime DOUGADOS (Paris)
Mr le Professeur Francis GUILLEMIN- Directeur de Thèse (Nancy)
Mr le Professeur Bernard MAZIERES- Rapporteur (Toulouse)
Mr le Professeur Philippe RAVAUD- Directeur de Thèse (Paris)

RAPPORTEURS : Mr le Professeur Bernard MAZIERES (Toulouse)
Mr le Professeur Gilles CHATELLIER (Paris)



A ma famille, qui est mon principal soutien.
Qu'elle trouve ici le témoignage de ma profonde affection.

A Monsieur le Professeur Philippe RAVAUD,

Pour la confiance qu'il m'a faite en dirigeant ce travail.

Son aide constante et efficace, ses encouragements et ses qualités humaines m'ont
été bénéfiques pour faire aboutir ce travail.

Qu'il trouve ici le témoignage de mon amitié sincère et entière, ainsi que ma
reconnaissance.

A Monsieur le Professeur Maxime DOUGADOS,
Qui m'a accueilli avec confiance dans son service.
Qu'il trouve ici le témoignage de mes sincères remerciements.



A monsieur Bruno GIRAUDEAU,
pour sa disponibilité et son aide efficace dans ce travail.
En témoignage de ma sincère reconnaissance.

A messieurs les Professeurs Jean-Luc DRAPE et Benoît ROUSSELIN,

Pour leur collaboration dans ce travail.

En témoignage de ma sincère reconnaissance.

A Monsieur le Professeur Serge BRIANÇON,

Qui m'a fait l'honneur d'accepter de participer à ce jury.

En témoignage de mes sentiments respectueux et de ma sincère considération.

A Monsieur le Professeur Bernard MAZIERES,

Qui m'a fait l'honneur d'accepter d'être mon rapporteur et de participer à ce jury.

Qu'il trouve ici le témoignage de ma profonde considération et de mes sincères
remerciements.

A Monsieur le Professeur Gilles CHATELLIER,

Qui m'a fait l'honneur d'accepter d'être mon rapporteur.

En témoignage de ma profonde considération et de mes sincères remerciements.

A Monsieur le Professeur Francis GUILLEMIN,
qui m'a fait l'honneur de participer à ce jury et dont les conseils m'ont été utiles.
Qu'il soit assuré de mes sincères remerciements et mon profond respect.

A Madame le Professeur Isabelle CHARY-VALCKENAERE,

Qui m'a fait l'honneur d'accepter de participer à ce jury.

Qu'elle trouve ici le témoignage de mon profond respect et de mes sincères
remerciements.

SOMMAIRE

I/	INTRODUCTION	Page 17
II/	TECHNIQUES D'IMAGERIE DANS LA COXARTHROSE	Page 22
	2-1) Radiographie standard	Page 23
	2-2) Autres techniques	Page 24
	2-2-1/ Radiographie micro focale	Page 24
	2-2-2/ Imagerie par résonance nucléaire	Page 25
	2-2-3/ Arthroscopie	Page 26
III/	MESURES DE LA PROGRESSION RADIOLOGIQUE DE LA COXARTHROSE	Page 28
	3-1) Intérêt de la mesure de l'interligne articulaire	Page 29
	3-2) Mesures qualitatives	Page 30
	3-2-1/ Score de Kellgren et Lawrence	Page 30
	3-2-2/ Critères individuels	Page 30
	3-3) Sources de variabilité	Page 31
	3-3-1/ Technique d'acquisition des radiographies	Page 31
	3-3-2/ Position du sujet et des membres inférieurs	Page 32

3-3-3/ Techniques de Mesure	Page 34
3-4) Intérêt de la standardisation	Page 39
IV/ REPRODUCTIBILITE	Page 41
4-1) Définition et intérêts	Page 42
4-2) Mesures statistiques de la reproductibilité	Page 44
4-2-1/ Variable continue	Page 44
4-2-2/ Variable qualitative	Page 49
V/ VARIABILITE DE LA MESURE DE L'INTERLIGNE DE L'INTERLIGNE ARTICULAIRE	Page 56
5-1) Etude de la concordance entre les mesures de l'interligne articulaire effectuées sur des radiographies réalisées en position debout et couchée	Page 57
5-1-1/ Etat des connaissances et objectifs de la recherche	Page 57
5-1-2/ Article 1	Page 59
Osteoarthritis of the hip: agreement between joint space width measurements on standing and supine conventional radiographs. G-R Auleley, B Rousselin, X Ayrat, R Edouard-Noel, M Dougados, P Ravaud <i>Ann Rheum Dis</i> 1998; 57: 519-523	

5-2) Etude de l'influence de la rotation des membres inférieurs et du centrage du rayon incident	Page 65
5-2-1/ Etat des connaissances et objectifs de la recherche	Page 65
5-2-2/ Article 2	Page 67
Measurement of joint space width in hip osteoarthritis: Influence of joint positioning and radiographic procedure. G-R Auleley, A Duche, J-L Drape, M Dougados, P Ravaud <i>Rheumatology In Press</i>	
5-3) Etude de l'influence de l'ordre de lecture des radiographies et du marquage du site de mesure dans les études longitudinales.	Page 88
5-3-1/ Etat des connaissances et objectifs de la recherche	Page 88
5-3-2/ Article 3	Page 91
Radiographic assessment of hip osteoarthritis progression: Impact of reading procedures for longitudinal studies. G-R Auleley, B Giraudeau, M Dougados, P Ravaud <i>Ann Rheum Dis 2000; 59: 422-427.</i>	
5-4) Essais thérapeutiques dans la coxarthrose : étude de l'influence du critère de jugement retenu et de la durée d'étude sur l'effectif à inclure.	Page 98
5-4-1/ Etat des connaissances et objectifs de la recherche	Page 98

5-4-2/ Article 4

Page 101

Sample size calculation according to study duration for different
Outcome measures in hip osteoarthritis clinical trials.

G-R Auleley, B Giraudeau, J-F Maillefert, M Lequesne,

M Dougados, P Ravaud

Soumis à publication.

VI/ CONCLUSION

Page 121

BIBLIOGRAPHIE

CHAPITRE I

INTRODUCTION

L'arthrose est une affection hétérogène qui associe des symptômes cliniques tels qu'une douleur articulaire et des images radiologiques telles que le pincement de l'interligne articulaire, des ostéophytes et une sclérose de l'os sous-chondral (*Altman 1986*). Il s'agit de la plus fréquente des maladies articulaires (*Kelsey 1988*). L'articulation de la hanche est l'une des articulations les plus fréquemment atteintes par l'arthrose (*Preidler 1996*). La prévalence radiologique de la coxarthrose est supérieure à 3% chez les sujets âgés de plus de 55 ans dans les populations caucasiennes (*Petersson 1996*) (*Felson 1988*). Elle augmente avec l'âge, passant de moins de 1% avant l'âge de 55 ans à plus de 10% à l'âge de 85 ans (*Danielsson 1997*). Dans les populations caucasiennes d'Europe du nord, une atteinte radiologique bilatérale a été observée chez 44% des sujets ayant une coxarthrose (*Jorring 1980*). En France, elle touche 2 à 4% de la population entre 40 et 70 ans (*Conrozier 1996*).

La coxarthrose est une des principales causes de handicap chez le sujet âgé (*Tepper 1993*) (*Fox 1996*) (*Theiler 1996*), la mise en place d'une prothèse de hanche pouvant se faire dans 53% des cas de coxarthrose (*Ledingham 1993*). On estime que tous les ans, 70 000 à 80 000 prothèses totales de hanche sont posées pour coxarthrose en France (*PMSE*). Aux Etats Unis, 75 prothèses totales de hanche pour 100 000 habitants et par an sont posées pour coxarthrose (*Hoaglund 1995*).

La radiographie standard permet de visualiser les aspects pathologiques de la maladie (*Resnick 1988*). Il s'agit d'un examen facilement réalisable, sans risque, coût-efficace pour diagnostiquer, quantifier la sévérité et suivre la progression de l'arthrose (*Altman 1991*) (*Jewell 1998*). Grâce au stockage des films, elle permet de réaliser aussi bien des études prospectives que rétrospectives.

Une approche quantitative permettant de décrire les images radiographiques de l'arthrose a été initialement développée par Kellgren et Lawrence en 1957 (*Kellgren et Lawrence 1957*). Actuellement, différentes méthodes sont disponibles pour quantifier les modifications structurales de l'arthrose visualisées par la radiographie (*Ravaud 1997*).

La destruction du cartilage articulaire est une des modifications structurales observées dans l'arthrose (*Resnick 1988*). L'interligne articulaire est considéré comme le reflet de l'épaisseur du cartilage articulaire (*Buckland-Wright 1995a*). Ainsi, le pincement de l'interligne articulaire est un des principaux signes radiologiques de l'arthrose.

La mesure de l'interligne articulaire est donc importante dans le diagnostic et l'évaluation de la progression de l'arthrose, aussi bien dans les essais thérapeutiques des

traitements anti-arthrosiques à visée structurale que dans les études épidémiologiques (Lequesne 1994) (Dieppe 1995) (Altman 1996) (GREGES 1996) (Hochberg 1997).

La progression de l'arthrose est considérée comme lente et le pincement de l'interligne articulaire varie de 0,1 à 0,6 mm par an (Dougados 1992) (Kirwan 1992) (Lequesne 1994b) (Goker 2000). L'évaluation de la progression au genou a montré que pour mettre en évidence une progression si faible, il est essentiel de réduire les sources de variabilité de la mesure de l'interligne articulaire (Leach 1970) (Lynch 1993) (Buckland-Wright 1994) (Buckland-Wright 1995) (Ravaud 1996 a) (Ravaud 1996 b) (Ravaud 1996 c) (Ravaud 1997).

A la hanche, de nombreuses questions concernant l'évaluation de la mesure de l'interligne articulaire restent néanmoins posées :

- Les facteurs de variabilité sont nombreux (rotation des pieds, centrage du rayon, procédures de lecture). Ainsi, les protocoles et les recommandations disponibles diffèrent sur la position du sujet lors de la réalisation de la radiographie de hanche ou du bassin (Resnick 1988) (Dieppe 1994) (Dieppe 1995) (Altman 1996) (Hochberg 1997). L'étude de l'influence de la position du patient (couchée ou debout) pendant la réalisation des radiographies a donné des résultats contradictoires (Evison 1987) (Teshima 1992) (Hanson 1993) (Conrozier 1997) (Pessis 1999). Par ailleurs, la rotation des membres inférieurs du patient pendant la réalisation des radiographies varie de 0° à 25° selon les études (Evison 1987) (Teshima 1992) (Hanson 1993) (Conrozier 1997) (Pessis 1999). L'influence de ce facteur sur l'évaluation de l'interligne est fréquemment négligée (Buckland-Wright 1998).
- Le type de radiographie à utiliser pour évaluer la coxarthrose n'est pas standardisé. La coxarthrose est fréquemment évaluée sur d'autres incidences radiologiques (urographie intraveineuse, lavement baryté en double contraste,...) (Croft 1990) (Ingarsson 2000).
- Les techniques de lecture varient elles aussi selon les études. L'évaluation de la progression de l'arthrose peut se faire en groupant ou non les radiographies d'un même patient (Altman 1987) (Ledingham 1993) (Dieppe 1994) (Speake 1995) (Dougados 1996) (Ravaud 1998). Par ailleurs, pour évaluer la progression radiologique de l'arthrose, certains auteurs utilisent le marquage du site de mesure sur l'interligne articulaire (Altman 1987) (Ravaud 1998), d'autres non (Dieppe 1997). Ces différents aspects sont rarement abordés par les protocoles et les recommandations disponibles (Dieppe 1994) (Dieppe 1995) (Altman 1996) (Hochberg 1997) (Resnick 1988). Enfin, leur

influence sur l'estimation du nombre de sujets nécessaires à inclure dans les études longitudinales n'a pas été encore évaluée.

- Le calcul du nombre de sujets nécessaires à inclure dans les essais thérapeutiques des traitements anti-arthrosiques à visée structurale dans la coxarthrose se heurte aux questions du critère de jugement à utiliser et du traitement des données manquantes.

Nous avons donc tenté de répondre à ces différentes questions.

Dans une première étude, nous avons évalué l'effet de la position du sujet sur la mesure de la hauteur de l'interligne articulaire de hanche. Nous avons mesuré la concordance entre les mesures de la hauteur de l'interligne articulaire sur des radiographies du bassin réalisées en position debout et couchée, chez des sujets atteints ou non de coxarthrose. Cette étude nous a permis de proposer des limites d'agrément entre ces mesures.

Dans une deuxième étude, nous avons évalué l'influence de la rotation interne des membres inférieurs et du centrage du rayon incident sur les mesures de la hauteur de l'interligne articulaire de hanche. Nous avons donc comparé les mesures de l'interligne articulaire sur des radiographies du bassin réalisées chez des sujets coxarthrosiques, en position couchée : les membres inférieurs en rotation interne de 15° puis de 5° dans un premier groupe de patients, le rayon incident centré sur l'ombilic puis sur le bord supérieur de la symphyse pubienne dans un deuxième groupe de patients. Cette étude nous a permis de proposer des recommandations concernant la rotation des membres inférieurs et le centrage du rayon incident afin de standardiser l'acquisition des radiographies standards du bassin.

Enfin, nous avons évalué l'effet de la standardisation de l'acquisition des radiographies sur les mesures de la hauteur de l'interligne articulaire. Pour cela, nous avons évalué la reproductibilité des mesures de la hauteur de l'interligne articulaire sur les radiographies standardisées du bassin, réalisées deux fois dans un intervalle de temps de court chez des sujets coxarthrosiques. Cette étude nous a permis de proposer une définition de la progression radiologique de la coxarthrose basée sur la reproductibilité des mesures de la hauteur de l'interligne articulaire (erreur de mesure).

Dans une troisième étude, nous avons évalué l'effet de la technique de lecture des radiographies du bassin dans les études longitudinales de la coxarthrose. Nous avons donc comparé différentes techniques de lecture (radiographies groupées par patient et lues dans l'ordre chronologique; radiographies groupées par patient et lues en aveugle de la chronologie; radiographies lues chacune isolément, en aveugle de la chronologie). Cette étude nous a permis de proposer une technique de lecture pour évaluer la progression de la coxarthrose.

Dans une dernière étude, nous avons estimé le nombre de sujets nécessaires à inclure dans les essais thérapeutiques de traitements anti-arthrosiques à visée structurale, en fonction de la durée des essais, en utilisant différents critères et prenant en compte les données manquantes. Cette étude nous a permis de montrer l'influence du traitement des données manquantes sur la variation du nombre de personnes-années de suivi.

CHAPITRE II

TECHNIQUES D'IMAGERIE DANS LA COXARTHROSE

2-1)- RADIOGRAPHIE STANDARD

La radiographie standard est actuellement la méthode la plus simple pour diagnostiquer et évaluer la progression des modifications structurales de l'arthrose (*Resnick 1988*) (*Altman 1996*) (*GREES 1997*) (*Hochberg 1997*). Elle permet de visualiser le pincement de l'interligne articulaire qui reflète la destruction cartilagineuse, de visualiser la sclérose sous-chondrale et les ostéophytes (*Hermodsson 1970*) (*Resnick 1988*) (*Preidler 1996*). La radiographie antéro-postérieure du bassin, de face, en position couchée est l'incidence la plus fréquemment réalisée (*Resnick 1988*).

Par ailleurs, les institutions internationales (World Health Organization and International League against Rheumatism (WHO/ILAR), Group for the respect of Ethics and Excellence in Sciences (GREES), Osteoarthritis Research Society (OARS) recommandent la mesure de l'interligne articulaire sur les radiographies standards comme critère de jugement principal dans les essais thérapeutiques évaluant les traitements anti-arthrosiques à visée structurale (*Lequesne 1994a*) (*Altman 1996*) (*GREES 1997*) (*Hochberg 1997*).

La radiographie standard a de nombreux avantages (*Adams 1991*) (*Buckland-Wright 1994*) (*Ayral 1998*). Il s'agit d'un examen simple, peu coûteux, facilement interprétable et bien admis. Il permet une appréciation cumulative de l'ensemble du processus arthrosique à la date de l'examen. Par ailleurs, les radiographies sont des documents objectifs permanents qui peuvent être réévalués. Les clichés radiographiques successifs peuvent être lus en aveugle de l'ordre chronologique et de l'identité du patient dans le cadre des études méthodologiques d'évaluation, des essais thérapeutiques ou des études épidémiologiques.

Cependant, la radiographie standard reste limitée par la quasi absence d'agrandissement radiographique due à la proximité du sujet avec la cassette contenant le film, la réduction du contraste de l'image et la mauvaise définition des bords osseux et des tissus mous (*Adams 1991*) (*Buckland-Wright 1994*). Par ailleurs, en l'absence d'image en trois dimensions, la superposition des structures antérieures peut compliquer l'interprétation des images (*Adams 1991*). Enfin, il existe une discordance entre les images radiographiques et la clinique (*Adams 1991*).

2-2)- AUTRES TECHNIQUES

2-2-1/ Radiographie micro focale

La radiographie micro focale ou radiographie à foyer fin a été proposée par Buckland-Wright (*Buckland-Wright 1989*). Elle se caractérise par l'utilisation d'une très petite source des rayons X inférieure à 15 microns de diamètre (contre 0,3 mm à 1 mm pour la radiographie standard). L'agrandissement de l'image obtenue (par 4, voire par 20) permet alors de voir des détails très fins, sans flou géométrique et d'améliorer la résolution spatiale (la taille des petites structures est de 25 à 50 microns) (*Buckland-Wright 1996*). Ces macroradiographies sont obtenues en plaçant l'articulation à 20-30 cm de la source et l'image est projetée sur la cassette placée à 1,5 m à 2 m. L'agrandissement de l'image obtenue est mesuré par le ratio de la distance entre la source et le film sur la distance entre la source et l'articulation. Cette technique a donc pour résultat de tripler la sensibilité de la mesure de l'interligne articulaire sur les macroradiographies (visualisant des images de 25 à 50 microns sur les macroradiographies, supérieures à 0,1 mm sur les radiographies standards).

Cependant, cette technique a une durée d'exposition longue, donc une irradiation plus forte pour le patient que la radiographie standard (*Buckland-Wright 1994*) (*Buckland-Wright 1996*). Elle n'est donc utilisée que pour des études chondrométriques spécifiques dans l'arthrose ou dans l'arthrite rhumatoïde (*Lequesne 1995*).

Pour Buckland-Wright et collaborateurs, la reproductibilité de la mesure automatique de l'interligne articulaire fémoro-tibial interne est meilleure sur des radiographies micro focales que sur des radiographies standards (coefficients de variation respectifs 1,6% et 3,2% sur les genoux non arthrosiques, 3,2% et 5,5% sur les genoux arthrosiques) (*Buckland-Wright 1995*). Dans cette étude, ces auteurs ont montré que la mesure de l'épaisseur du cartilage articulaire du genou, sur des clichés d'arthrographie de 20 patients arthrosiques, était corrélée à la mesure de la hauteur de l'interligne articulaire sur les radiographies micro focales en sémiflexion. Ils concluaient donc que la radiographie micro focale était une méthode valide et reproductible dans l'évaluation de la progression de l'arthrose.

Cette technique peut être utilisée pour évaluer la plupart des parties du corps (*Buckland-Wright 1998*) mais n'a toujours pas été évaluée dans la coxarthrose (*Buckland-Wright 1996*).

2-2-2/ Imagerie par résonance nucléaire

L'imagerie par résonance nucléaire (IRM) est une méthode non invasive et non irradiante permettant de visualiser la totalité de l'articulation (*Fernandez-Madrid 1994*) (*Hutton 1995*) (*Peterfy 1996*) (*Loeuille 1998*).

Elle permet une évaluation quantitative de la morphologie du cartilage. Cependant, la sensibilité des images dans l'articulation du genou peut varier de 20% à 96% et la spécificité de 58% à 95% selon la technique d'acquisition utilisée (*Recht 1993*). Actuellement, l'IRM tridimensionnelle (3D) avec séquences en écho de gradient et suppression de graisse est la technique de choix (*Recht 1993*) (*Disler 1995*) (*Peterfy 1996*) (*Recht 1996*). Dans des études récentes incluant plus de 40 patients, cette technique avait une sensibilité de 81% à 86% et une spécificité de 97% pour détecter des defects cartilagineux visibles à l'arthroscopie (*Disler 1995*) (*Recht 1996*). Cependant, Tyrrel a montré que sur des séquences en écho de gradient, l'IRM ne détectait que 36% des lésions cartilagineuses superficielles visualisées par l'arthroscopie (*Tyrrel 1988*). De plus, la profondeur des lésions visualisées par l'arthroscopie pouvait être sous estimée par l'IRM (*Handelberg 1990*) (*Blackburn 1994*) (*Broderick 1994*). Ainsi, sur des séquences en spin écho et en écho de gradient, l'IRM peut être normale dans 5% des cas de chondropathie sévère (*Blackburn 1994*). Enfin, la mesure du volume cartilagineux total par l'IRM n'est pas sensible aux modifications localisées puisqu'on manque d'information concernant la distribution des modifications cartilagineuses dans l'articulation (*Peterfy 1998*).

Cette technique peut être également utilisée pour évaluer la qualité du cartilage (contenu en eau du cartilage) (*Selby 1995*) (*Peterfy 1998*). Mais elle doit être encore validée et optimisée pour cette fonction (*Altman 1996*).

Bien que la hanche soit la plus grosse articulation touchée par l'arthrose, l'exploration de la hanche par l'IRM pose des difficultés particulières (*Hutton 1995*). Il s'agit d'une articulation très congruente permettant difficilement de distinguer les surfaces articulaires. L'absence relative de liquide articulaire dans la cavité synoviale étroite de la hanche rend difficile le tracé des surfaces. Enfin, les surfaces articulaires très incurvées, favorisent des effets de volume importants (partial-volume effects) dans tous les plans (*Hutton 1995*) (*Loeuille 1998*), sauf en cas d'utilisation d'une résolution spatiale élevée (*Loeuille 1998*) (*Peterfy 1998*).

Cependant, Li et collaborateurs ont montré que la gradation de la sévérité des lésions arthrosiques sur les images d'IRM corrélait avec la gradation de la sévérité radiologique sur la même échelle (*Li 1988*). Mais, la mesure de l'épaisseur cartilagineuse de la hanche en IRM est plutôt décevante. Une étude de 10 hanches en post-mortem a montré qu'il n'existait pas de

corrélation suffisante entre les mesures du cartilage sur des coupes anatomiques et sur les images d'IRM en spin écho et suppression de graisse (*Holder 1992*).

Les propriétés métrologiques de l'IRM dans l'arthrose ont été peu étudiées (*Peterfy 1994*) (*Pilch 1994*) (*Drapé 1998*). La reproductibilité de la mesure du volume cartilagineux dans le compartiment fémoro-tibial du genou sain en IRM 3D apparaît mauvaise puisqu'on peut observer jusqu'à 20% d'erreurs de mesure sur des objets de volume connu (*Pilch 1994*). Par contre, Drapé et collaborateurs, dans une étude de 43 patients arthrosiques, ont trouvé une corrélation significative entre des lésions cartilagineuses du compartiment fémoro-tibial interne étudiées en IRM 3D (séquences en écho de gradient pondéré en T1) et en arthroscopie ($r=0,83$) (*Drapé 1998*). Ils ont également trouvé une reproductibilité transversale intra-observateur et une reproductibilité transversale inter-observateur des mesures en IRM raisonnables (coefficients de corrélation intra-classe respectifs de 0,94 et 0,80). Cependant, dans une étude de 10 sujets, Jonsson et collaborateurs ont eux trouvés que la reproductibilité de la mesure de l'interligne articulaire sur les images d'IRM était moins bonne que celle obtenue sur les radiographies standards, avec des coefficients de variation respectifs de 11,9% et 6,5% pour le genou, de 14,4% et 6,34% pour la hanche (*Jonsson 1992*). Mais une évaluation de la reproductibilité longitudinale de cette technique est nécessaire.

L'IRM est un examen prometteur dans l'évaluation des études épidémiologiques et des essais thérapeutiques évaluant les traitements anti-arthrosiques à visée structurale dans la coxarthrose. Mais une amélioration technique de cet examen (résolution spatiale) est nécessaire (*Peterfy 1998*). Le développement d'IRM dédiée aux articulations devrait permettre de pallier les limites de son utilisation en raison de son coût et de sa disponibilité.

2-2-3/ Arthroscopie

L'arthroscopie est le seul examen offrant une vision directe, fine, et complète des surfaces articulaires. Elle permet d'évaluer la profondeur et l'étendue des lésions. L'enregistrement vidéo de l'examen permet de quantifier ultérieurement ces lésions (*Ayral 1993*) (*Dougados 1994*) (*Ayral 1996*).

Dans l'arthrose, cet examen a surtout été évalué au genou (*Ayral 1996*). La reproductibilité intra-observateur du score SFA de la Société Française d'Arthroscopie d'une part et celle de l'évaluation globale du cartilage mesurée sur une échelle visuelle analogique de 100 mm d'autre part, paraissent satisfaisantes (coefficients de corrélation intra-classe respectifs de 0,928 et 0,989).

L'arthroscopie de la hanche est beaucoup moins fréquemment pratiquée. Elle a de nombreux inconvénients. Elle peut nécessiter dans certains cas d'utiliser une force de traction importante pour introduire l'endoscope et la durée de l'examen est d'au moins 45 minutes (*Villar 1994*). Par ailleurs, elle ne peut être répétée facilement, en particulier dans le cadre d'études, et a un coût élevé (*Villar 1994*) (*Ayral 1998*). Son intérêt paraît donc limitée dans les études épidémiologiques et les essais thérapeutiques dans la coxarthrose.

D'autres techniques d'imagerie peuvent être utilisées dans l'arthrose (*Adams 1991*):

- la tomodensitométrie est moins performante que la radiographie pour mesurer l'interligne articulaire (*Conrozier 1996*),
- Les ultrasons à haute fréquence ont été peu évalués dans la mesure de la hauteur de l'interligne articulaire, en particulier dans la coxarthrose (*Jonsson 1992*) (*Sada 1994*) (*Castriota-Scanderberg 1996*). A la hanche, cet examen a une moins bonne reproductibilité que la radiographie standard et l'IRM (coefficients de variation de 18,9% contre 6,34% et 14,4%) (*Jonsson 1992*).

Au total, il apparaît que la radiographie standard est la technique d'imagerie la plus utilisée dans la coxarthrose. L'IRM, recommandée dans l'évaluation de la progression de l'arthrose (*Altman 1996*), reste un examen prometteur pour l'exploration de la coxarthrose.

CHAPITRE III

MESURES DE LA PROGRESSION RADIOLOGIQUE DE L'INTERLIGNE ARTICULAIRE DANS LA COXARTHROSE

3-1)- INTERÊT DE LA MESURE DE L'INTERLIGNE ARTICULAIRE

Le pincement de l'interligne articulaire est un signe important pour diagnostiquer l'arthrose de la hanche (*Altman 1991*) (*Hart 1998*).

A la hanche, la mesure de la hauteur de l'interligne articulaire au point le plus pincé a une meilleure reproductibilité transversale que d'autres critères radiologiques, en particulier la taille de l'ostéophyte (reproductibilité intra-observateur: coefficients kappa de 0,81 contre 0,44; reproductibilité inter-observateur : coefficients kappa de 0,70 contre 0,33) (*Croft 1990*).

Ce critère semble meilleur que les autres critères radiologiques pour déterminer la progression radiologique de la maladie (*Altman 1987*). Par ailleurs, Dougados et collaborateurs ont montré que la mesure de l'interligne articulaire est plus sensible que la score de Kellgren et Lawrence pour suivre l'évolution de la coxarthrose (*Dougados 1995*). Dans une autre étude mesurant la hauteur de l'interligne articulaire sur des radiographies standards du bassin de 30 sujets coxarthrosiques, cette même équipe a montré que la reproductibilité longitudinale à 1 an de ce critère est acceptable (coefficients de corrélation intra-classe de 0,77 pour la reproductibilité intra-observateur et de 0,68 pour la reproductibilité inter-observateur) (*Dougados 1996a*).

Par ailleurs, la mesure de la hauteur de l'interligne articulaire à un moment donné et sa variation à 1 an semblent préjuger de l'état de morbidité à long terme des patients coxarthrosiques (*Dougados 1999*) (*Conrozier 1998*). Les patients ayant un interligne de moins de 1,5 mm à un moment donné auraient un risque 1,8 fois plus important de recourir à la mise en place d'une prothèse totale de hanche dans les trois années suivantes que ceux ayant un interligne plus important (*Dougados 1999*).

Enfin, ce critère est actuellement le principal critère de jugement recommandé dans les études épidémiologiques et les essais thérapeutiques évaluant les traitements anti-arthrosiques à visée structurale (*Lequesne 1994*) (*Dieppe 1995*) (*Altman 1996*) (*GRÈES 1996*) (*Hochberg 1997*).

3-2)- METHODES QUALITATIVES

3-2-1/ Score de Kellgren et Lawrence

Le score de Kellgren et Lawrence a été initialement proposé par ces deux auteurs en 1957 (*Kellgren 1957*). Il s'agit d'un score ordinal en quatre stades, évaluant la sévérité des atteintes radiologiques de l'arthrose, prenant en compte l'existence d'ostéophytes, d'un pincement du cartilage articulaire et de celle d'une sclérose sous-chondrale. Classiquement, le diagnostic d'arthrose dans les études épidémiologiques reposait sur les modifications radiologiques caractéristiques décrites par Kellgren et Lawrence en 1957.

Récemment, de nombreux auteurs ont noté des limites à l'utilisation de ce score (*Spector 1994a*) (*Spector 1994b*) (*Murphy 1995*). Premièrement, son interprétation n'est pas précise: la description des grades varie selon les auteurs, aboutissant à des classifications différentes dans les études épidémiologiques par des équipes qui, toutes, se réclament de la même référence (*Hart 1995*) (*Ravaud 1995*). Par ailleurs, l'importance que ce score accorde à la présence d'ostéophytes pour la classification de l'arthrose a été également critiquée (*Wood 1976*) (*Brandt 1991*). Ainsi, l'association d'un pincement de l'interligne articulaire, d'une sclérose sous-chondrale et d'une déformation ne serait pas classée comme arthrose en l'absence d'ostéophytes (*Kallman 1989*) (*Hart 1995*). Enfin, le score de Kellgren et Lawrence qui n'a pas été initialement décrit pour une évaluation longitudinale, n'apparaît pas sensible à la progression de l'arthrose (*Spector 1994b*) (*Murphy 1995*) (*Ravaud 1996b*).

3-2-2/ Critères individuels

Des échelles spécifiques pour chaque articulation et accordant plus d'importance au pincement articulaire (*Croft 1990*) ou au pincement articulaire et aux ostéophytes (*Lane 1993*) ont été développés afin d'éviter les problèmes posés par l'utilisation du score de Kellgren et Lawrence. Elles permettent d'évaluer individuellement les ostéophytes, le pincement articulaire, les géodes et la sclérose sous-chondrale. Pour la hanche, ces échelles comportent 4 (*Lane 1993*) ou 6 grades (*Croft 1990*).

Un atlas illustrant les différents grades peut être utilisé pour évaluer individuellement ces différents critères (*Altman 1995*).

3-3)- SOURCES DE VARIABILITE

De nombreux auteurs ont souligné l'importance de la reproductibilité de la mesure de l'interligne articulaire dans l'évaluation de la progression de l'arthrose (*Buckland-Wright 1994*) (*Lequesne 1994b*) (*Ravaud 1997*). Les sources de variabilité pouvant compromettre la reproductibilité de cette mesure ont été décrites (*Buckland-Wright 1994*) (*Ravaud 1997*). Leurs effets sur la mesure de l'interligne articulaire sont bien connus au genou et ont été peu évalués à la hanche.

3-3-1- Technique d'acquisition des radiographies

Trois facteurs de la technique d'acquisition des radiographies peuvent compromettre la reproductibilité de la mesure de l'interligne articulaire : la distance entre la source des rayons X et le centre de l'articulation étudiée, la distance entre le centre de l'articulation étudiée et la plaque radiographique, et l'inclinaison ou le centrage du rayon incident.

** Distances entre source des rayons X et articulation, entre articulation et plaque*

L'augmentation de la distance entre le centre de l'articulation étudiée et la plaque radiographique peut provoquer un agrandissement de l'image et donc une erreur dans la mesure si un facteur de correction n'est pas utilisé (*Buckland-Wright 1994*). Par ailleurs, l'agrandissement peut être associé à des images floues en raison d'une perte de définition radiographique (*Buckland-Wright 1994*). La distance habituellement conseillée entre la source des rayons X et le centre de l'articulation étudiée est de 100 cm (*Ravaud 1997*). Cependant, une distance de 110 cm est parfois utilisée pour radiographier la hanche (*Turula 1985*) (*Hansson 1993*) (*Conrozier 1997*). L'effet de ce facteur sur la mesure de la hauteur de l'interligne de la hanche n'a pas été encore étudié.

** Centrage et angle du rayon incident*

Les modifications de l'inclinaison du rayon incident pendant la réalisation des radiographies peuvent entraîner une superposition des extrémités osseuses et une modification de l'aspect de l'interligne articulaire, le rayon incident n'étant plus ni parallèle aux surfaces articulaires ni perpendiculaire à la plaque radiographique (*Buckland-Wright 1994*). Dans une étude de radiographies du genou, réalisées avec une inclinaison du rayon incident de 5° et de 10° vers le bas par rapport au centre du genou, chez 10 sujets volontaires non arthrosiques,

Ravaud et collaborateurs ont récemment montré que la hauteur de l'interligne articulaire diminuait respectivement de 0,95 mm et 1,13 mm dans le compartiment fémoro-tibial interne (Ravaud 1996a).

Par ailleurs, la modification de la mesure de l'interligne articulaire paraît non significative sur des radiographies du genou réalisées en extension chez des volontaires sains, avec déplacement du rayon incident de sa position initiale par rapport au centre de l'articulation (Ravaud 1996a). La hauteur de l'interligne articulaire du compartiment fémoro-tibial interne ne diminue respectivement que 0,05 mm et 0,15 mm avec un centrage du rayon incident à 0,5 cm et 1 cm au dessous du centre de l'articulation du genou. Cependant, cette diminution peut représenter 17% de l'interligne articulaire mesuré sur des radiographies réalisées avec le rayon centré sur le centre de la rotule (Fife 1991). Lynch et collaborateurs ont observé une modification de 0,02 mm de l'interligne articulaire pour 1 mm de déplacement vertical du rayon central du faisceau des rayons X par rapport au centre du genou pendant la réalisation de macroradiographies de 10 genoux de cadavres (Lynch 1993).

A la hanche, c'est la radiographie standard de face du bassin qui est la plus utilisée. Dans cette radiographie, le rayon incident est centré au dessus du bord supérieur de la symphyse pubienne, la hanche se trouve alors dans une zone relativement centrale du cliché. D'après certains auteurs, ce cliché pourrait être responsable d'une modification de l'interligne articulaire puisque le rayon incident n'est pas centré sur la hanche (Buckland-Wright 1998). Ainsi, certains auteurs mesurent l'interligne articulaire sur des radiographies centrées sur la tête fémorale (Conrozier 1997). D'autres auteurs utilisent des radiographies réalisées avec un centrage situé environ 10 cm au dessus du bord supérieur de la symphyse pubienne (Croft 1990). Cependant, l'influence du centrage du rayon incident sur la mesure de l'interligne articulaire n'a pas encore été réellement évaluée à la hanche.

3-3-2/ Position du sujet et des membres inférieurs

La position debout ou couchée et le degré de rotation des membres inférieurs sont les principales sources de variabilité concernant la position du sujet pendant la réalisation de la radiographie standard.

*** Mise en charge**

L'influence de la mise en charge est bien connue au genou. Plusieurs auteurs ont observé une différence entre les mesures de l'interligne articulaire obtenues sur les radiographies du genou réalisées en charge et en décharge (Leach 1970) (Thomas 1975). Leach et

collaborateurs ont observé que chez 33% des 130 patients ayant des douleurs du genou, la différence de hauteur de l'interligne articulaire mesurée sur des radiographies réalisées en charge et en décharge était égale ou supérieure à 2 mm (*Leach 1970*). Cette différence atteignait 5 mm chez certains patients. Ces discordances ont conduit à recommander de réaliser les radiographies standards du genou ou de la hanche en charge.

Dans une étude de 151 hanches (108 patients) dont 63 avaient un pincement articulaire, Hansson et collaborateurs ont mesuré la hauteur de l'interligne articulaire au point médian, avec une règle standard, sur des radiographies standards de la hanche, réalisées en position debout et couchée (*Hansson 1993*). La hauteur moyenne de l'interligne articulaire était de $4,2 \pm 1,7$ mm sur les radiographies réalisées en position couchée et de $3,9 \pm 1,7$ mm sur les radiographies réalisées en charge ($p < 0,0001$).

Conrozier et collaborateurs ont eux aussi comparé l'interligne articulaire mesurée sur des radiographies standards du bassin (57 hanches dont 46 avec coxarthrose) et de la hanche (56 hanches avec coxarthrose) réalisées en position debout et couchée (*Conrozier 1997*). Dans le sous-groupe de hanches avec un interligne inférieur ou égal 2,5 mm, la hauteur de l'interligne articulaire au point le plus pincé, mesurée grâce à un logiciel d'analyse d'image sur des radiographies digitalisées, était 11,5% en moyenne plus petite sur les radiographies de hanche réalisées en position debout par rapport à celles réalisées en position couchée (6,7% sur les radiographies du bassin).

Cependant, dans une série de 38 patients (13 cas de coxarthrose), seuls 5 patients avaient un interligne plus petit de 0,1 à 0,5 mm (les autres étaient identiques) sur les radiographies réalisées en position debout par rapport à celles réalisées en position couchée (*Evison 1987*).

A l'inverse, Teshima et collaborateurs ont comparé la surface articulaire des deux hanches de 20 patients amputés de l'extrémité distale d'une des jambes (décharge depuis 5 à 92 mois) (*Teshima 1992*). Ces auteurs ont trouvé que dans la moitié des cas, la surface articulaire du côté amputé était inférieure d'au moins 10% à celle du côté controlatéral (d'au moins 30% chez 6 patients).

* Position des membres inférieurs

L'influence de la rotation des membres inférieurs pendant la réalisation des radiographies standards a été étudiée au genou. En modifiant la rotation des membres inférieurs de 5° à 15° par rapport à la position idéale, Lynch et collaborateurs ont conclu que ce facteur ne modifiait pas les mesures de l'interligne articulaire sur des macroradiographies

de 3 genoux de cadavre (*Lynch 1993*). Cependant, ils observaient une augmentation de l'erreur de mesure de l'interligne articulaire de 0,15 mm pour 10° de modification de rotation. Plus récemment, Ravaud et collaborateurs ont montré que, par rapport à la position idéale (épines tibiales centrées sur l'échancrure condylienne), une rotation externe des membres inférieurs de 15° ou 30° pendant la réalisation des radiographies standards des genoux, provoquait une diminution significative de la hauteur de l'interligne articulaire dans le compartiment fémoro-tibial interne de 10% ou 18% (*Ravaud 1996a*).

A la hanche, la rotation des membres inférieurs peut varier de 0° à 25° pendant la réalisation des radiographies (*Hansson 1993*) (*Buckland-Wright 1999*). Cependant, l'influence de ce facteur sur la mesure de l'interligne articulaire a été rarement évaluée. Buckland-Wright a répété des radiographies chez 9 cadavres en modifiant la rotation des membres inférieurs (*Buckland-Wright 1998*). La hauteur de l'interligne articulaire au point le plus étroit était de 4,1±0,6 mm avec une rotation interne des membres inférieurs et de 4,6±0,6 mm sans rotation des membres inférieurs ($p<0,01$). Selon cet auteur, l'influence de la rotation des membres inférieurs sur la mesure de l'interligne articulaire pourrait être due à l'action puissante des muscles rotateurs de la hanche, entraînant la tête fémorale en dedans contre le cotyle (*Buckland-Wright 1998*).

3-3-3/ Techniques de mesure

La reproductibilité de la mesure de l'interligne articulaire peut être compromise par la technique choisie pour mesurer la progression de l'interligne.

*** Instrument**

La plupart des instruments utilisés pour mesurer l'interligne articulaire sont communs au genou et à la hanche. Ils ont été plus souvent évalués au genou qu'à la hanche. Il s'agit principalement de :

- la règle graduée en demi-millimètres
- le compas à pointe sèche
- la loupe graduée au 10^{ème} de mm
- la digitalisation des limites de l'interligne
- la lecture automatique d'images digitalisées

La mesure de l'interligne articulaire du genou arthrosique (hauteur moyenne égale à 4,9 ± 1,9 mm) avec la règle graduée a des coefficients de variation de 7,7% pour la reproductibilité intra-lecteur et de 10,8% pour la reproductibilité inter-lecteur (*Mazières 1990a*).

Cependant, la reproductibilité de cet instrument semble moins bonne pour les petits que pour les grands interlignes : les coefficients de variation peuvent varier de 19% à 26,4% pour des interlignes inférieurs 4 mm (*Glimet 1989*) (*Mazières 1990a*) alors qu'ils sont inférieurs à 10% pour de plus grand interlignes (*Mazières 1990a*). Dans une étude incluant 10 sujets, les coefficients de variation mesurant la reproductibilité intra-lecteur de la mesure de la hauteur de l'interligne articulaire avec une règle demi-millimétrée, sur des radiographies standards réalisées en position debout et couchée, étaient respectivement de 6,5% et 5,44% au genou et de 6,34% et 6,32% à la hanche (*Jonsson 1992*).

La reproductibilité de la mesure de l'interligne articulaire avec un compas à vis stabilisatrice, suivie d'une lecture entre les pointes sur une règle, a été évaluée dans une série de 64 radiographies du genou (*Massé 1990*). Les coefficients de variation variaient de 9% à 17% pour la reproductibilité inter-lecteur. A la hanche, sa reproductibilité inter-lecteur est proche, avec un coefficient de variation variant de 9% à 10% (*Lequesne 1994b*).

La reproductibilité intra-lecteur de la loupe graduée au 10^{ème} de mm a été récemment évaluée à la hanche par Lequesne et collaborateurs dans une étude de 26 radiographies standards du bassin de patients souffrant d'une coxarthrose (hauteur de l'interligne variant de 0,3 à 4,2 mm) (*Lequesne 1999*). A 3 mois d'intervalle, un lecteur unique répétait la mesure de la hauteur de l'interligne articulaire au point le plus pincé, avec ou sans marquage du site de mesure. Les coefficients de variation (coefficients de corrélation intra-classe) étaient de 5,7% à 8% (0,97 à 0,98) avec marquage du site de mesure, et de 15,3% (0,88) sans marquage du site de mesure. Dougados et collaborateurs ont évalué la reproductibilité de la loupe graduée au 10^{ème} de mm en mesurant la hauteur de l'interligne au point le plus pincé sur des radiographies répétées à 1 an d'intervalle (*Dougados 1996a*). Dans cette étude de 30 couples de radiographies de patients souffrant de coxarthrose, la reproductibilité était étudiée en utilisant la méthode de Bland et Altman définissant les limites d'agrément entre deux mesures à 1,96 SD des différences entre les mesures (SD : déviation standard) et les coefficients de corrélation intra-classe. Deux lecteurs étaient mis à contribution. Les limites d'agrément (coefficient de corrélation intra-classe) étaient de 0,80 mm (0,85) pour la reproductibilité intra-lecteur transversale et de 1,04 mm (0,78) pour la reproductibilité inter-lecteur transversale. Ils étaient de 0,56 mm (0,77) pour la reproductibilité intra-lecteur longitudinale, et de 0,68 (0,68) pour la reproductibilité inter-lecteur longitudinale.

Conrozier et collaborateurs ont évalué la reproductibilité de la mesure de l'interligne articulaire de la hanche après digitalisation du cliché radiographique dans deux études (*Conrozier 1990*) (*Conrozier 1995*). Leur technique de mesure consiste à digitaliser la

radiographie puis à dessiner les contours de la surface de l'interligne articulaire sur l'image ainsi digitalisée. Les mesures sont alors automatiques pour la surface de l'interligne articulaire et l'épaisseur articulaire moyenne, définie comme le rapport de la surface de l'interligne articulaire sur la longueur du segment de l'interligne étudié (*Conrozier 1990*). Dans une étude de 22 interlignes, ces deux paramètres étaient mesurés cinq fois chacun par deux lecteurs (*Conrozier 1990*). Les coefficients de variation des mesures de ces deux paramètres étaient respectivement de 5,6% et 3,8%.

Dans la deuxième étude incluant 6 hanches (3 hanches ayant une coxarthrose), chaque interligne articulaire était mesuré dix fois (*Conrozier 1995*). Les coefficients de variation intra-lecteur étaient de 1,7% pour la surface articulaire moyenne, 2,9% pour la hauteur moyenne de l'interligne au point le plus pincé et de 1,2% pour l'épaisseur moyenne de l'interligne.

Ces auteurs ont aussi évalué la reproductibilité de l'ensemble du processus radiographique et de la mesure de l'interligne articulaire sur 20 hanches (10 arthrosiques) (*Conrozier 1993*). Chaque articulation était indépendamment radiographiée par trois radiologues différents. L'interligne était ensuite mesurée 5 fois en aveugle par deux lecteurs entraînés. Les coefficients de variation inter-lecteur étaient de 3,306% pour l'épaisseur moyenne et de 4,32% pour la surface articulaire. Les coefficients de variation intra-lecteur étaient de 2,97% et de 3,54% pour l'épaisseur moyenne et de 4,32% et 5,13% pour la surface articulaire.

Dougados et collaborateurs ont comparé les mesures de l'interligne articulaire avec un compas à point sèche et report sur une règle demi-millimétrée avec celles de l'interligne articulaire après digitalisation du cliché radiographique du genou (technique similaire à celle utilisée par Conrozier et collaborateurs à la hanche) (*Dougados 1991*). Les coefficients de variation intra-lecteur étaient de 3% pour les mesures de l'interligne avec compas et de 4,4% pour les mesures de surface articulaire après digitalisation de l'image. Le coefficients de variation inter-lecteur était de 11,8% pour la mesure de la surface articulaire.

D'autres techniques de mesure automatique de l'interligne articulaire sur des images digitalisées ont été développées et évaluées au genou (*Dacre 1989a*) (*Dacre 1989b*) (*Lynch 1993*) (*Buckland-Wright 1994*). Elles pourraient avoir une reproductibilité acceptable avec des coefficients de variation inférieurs à 2%. Cependant, ces différentes techniques de mesure automatique de l'interligne articulaire n'ont pas encore été comparées (*Buckland-Wright 1999*).

Récemment, Ravaud et collaborateurs ont comparé la reproductibilité des mesures de la hauteur de l'interligne articulaire dans le compartiment fémoro-tibial interne de 30 genoux de patients souffrant d'arthrose de quatre instruments de mesure : règle graduée en demi-millimètres, compas à pointe sèche avec report de la mesure sur la règle, loupe graduée au

10^{ème} de mm et mesure automatique après digitalisation des bords de l'interligne (*Ravaud 1996c*). La reproductibilité intra et inter-observateur (3 lecteurs) de tous ces instruments était élevée (coefficients de corrélation intra-classe variant respectivement de 0,95 à 0,98 et de 0,91 à 0,97).

* Site ou région de mesure

A la hanche, l'interligne articulaire est mesuré pour la hauteur, la surface et l'épaisseur moyenne (*Conrozier 1990*) (*Conrozier 1993*) (*Conrozier 1995*). Ces différentes mesures peuvent être obtenues par des méthodes automatiques (*Teshima 1992*) (*Conrozier 1995*) (*Ravaud 1996c*) ou, plus fréquemment pour la hauteur de l'interligne, par des méthodes manuelles (*Buckland-Wright 1999*).

La mesure de la surface articulaire paraît moins précise que celle de l'épaisseur moyenne de l'interligne articulaire (*Conrozier 1990*) (*Conrozier 1993*) (*Conrozier 1995*). Par ailleurs, elle nécessite une standardisation de la taille de l'articulation, qui diffère selon les individus et pourrait être influencée par la croissance des ostéophytes (*Buckland-Wright 1999*). La définition de la surface articulaire semble également varier selon les auteurs (*Teshima 1992*) (*Conrozier 1993*).

La mesure de la surface de l'interligne articulaire ne semble pas meilleure que celle de la hauteur de l'interligne articulaire pour évaluer la progression de la coxarthrose. Dans une étude de 22 patients souffrant de coxarthrose suivis pendant 10 à 15 mois, sa sensibilité au changement était de 0,32, alors que celle de la loupe graduée au 10^{ème} de mm était de 0,37 (*Dougados 1995*).

La mesure de l'épaisseur moyenne de l'interligne articulaire est définie à partir de celle de la surface articulaire. Elle a donc les mêmes limites que la mesure de la surface articulaire. Par ailleurs, les mesures de l'épaisseur moyenne et de la surface articulaire doivent encore être validées comme critères de jugement (*Buckland-Wright 1999*).

A la hanche, la hauteur de l'interligne articulaire peut être mesurée au point le plus pincé (*Dougados 1995*) (*Dougados 1996*) ou au point médian (*Hansson 1993*) ou sur plusieurs sites dans la même étude (*Jonsson 1992*). L'influence respective de ces différents sites n'a pas encore été évaluée.

Cependant, la mesure de la hauteur de l'interligne articulaire au point le plus pincé est actuellement recommandée pour évaluer la progression radiologique à la hanche (*Altman 1996*).

Enfin, le choix des points anatomiques permettant de déterminer les limites de la région ou du site de mesure peut varier puisqu'il dépend du lecteur. L'influence de l'utilisation du marquage du site ou la région de mesure reste encore à évaluer.

* Lecteur

De nombreux auteurs ont montré que la reproductibilité de la mesure de l'interligne articulaire différait selon le niveau d'entraînement des lecteurs (*Dougados 1996a*) (*Ravaud 1996b*). Par ailleurs, Altman a montré que l'utilisation de plusieurs lecteurs pouvait améliorer la reproductibilité et la capacité des critères radiologiques qualitatifs à évaluer la progression de l'arthrose (*Altman 1987*). Cependant, l'influence de ce dernier facteur sur la mesure de l'interligne articulaire n'a pas encore été étudiée.

Il apparaît donc que de nombreux paramètres peuvent influencer la reproductibilité de la mesure de l'interligne articulaire à la hanche. Bien que la hanche soit une articulation fréquemment atteinte par l'arthrose, il n'existe pas à ce jour de consensus sur une méthode de mesure standardisée de l'interligne articulaire.

3-4) INTERÊT DE LA STANDARDISATION DE LA MESURE DE L'INTERLIGNE ARTICULAIRE

La progression de l'arthrose est considérée comme lente et les modifications de l'interligne articulaire peuvent être difficiles à mettre en évidence, même après de nombreuses années d'évolution (*Spector 1992*).

Plusieurs sources de variabilité peuvent affecter la mesure de l'interligne articulaire, depuis la réalisation des radiographies jusqu'à la mesure proprement dite (*Ravaud 1997*). Il apparaît donc essentiel d'améliorer la reproductibilité de la mesure de l'interligne articulaire (*Wright 1992*).

La standardisation de la mesure de l'interligne articulaire a été quelquefois tentée. Cependant, elle a plus souvent concerné certaines étapes isolées du processus (position du patient ; mesure de l'interligne) que l'ensemble du processus (*Sui 1991*) (*Jonsson 1992*). L'utilisation de la fluoroscopie pour reproduire la position de l'articulation (*Jonsson 1992*) augmente l'irradiation des patients. Par ailleurs, elle ne peut être utilisée partout. Enfin, associée à des recommandations, son utilisation ne paraît pas supérieure à l'utilisation des seules recommandations pour évaluer la progression de l'interligne articulaire (*Ravaud 1998*).

Certains auteurs ont utilisé un appareillage coûteux et lourd à manipuler pour immobiliser les membres inférieurs pendant les radiographies (*Sui 1991*) (*Jonsson 1992*). Cependant, l'amélioration de la reproductibilité de la mesure doit se faire sans augmenter les coûts ni générer des complications inutiles (*Sui 1991*) (*Wright 1992*). L'utilisation d'une feuille de papier sur laquelle l'empreinte des pieds est dessinée immédiatement après la réalisation de la radiographie a été recommandée pour réaliser des radiographies successives en position debout (*Altman 1996*). Cette méthode apparaît moins pratique pour les radiographies réalisées en position couchée.

D'autre part, les méthodes de lecture automatiques d'images digitalisées sont censées permettre des mesures plus précises de l'interligne articulaire que celles obtenues par les méthodes de lecture manuelles. Elles sont souvent utilisées sur des images obtenues sans contrôle des conditions techniques d'acquisition des radiographies ni de la position du patient pendant la réalisation des radiographies (*Teshima 1992*) (*Conrozier 1997*). Par ailleurs, il n'existe pas encore de consensus sur les méthodes de lecture automatique d'images digitalisées (*Buckland-Wright 1999*). Enfin, elles ont un certain coût d'investissement et leur supériorité sur

les méthodes de lecture manuelles telles que la loupe graduée au 10^{ème} de mm n'est pas définitivement établie (*Dougados 1991*) (*Dougados 1995*).

La standardisation de la mesure de l'interligne articulaire de la hanche doit donc être précédée d'une évaluation de l'influence des différentes sources de variabilité à toutes les étapes. Elle a pour but d'obtenir des résultats fiables pendant des examens successifs.

Elle permet aussi de définir l'erreur de mesure de l'ensemble du processus (position du patient pendant les radiographies, procédure technique d'acquisition des radiographies et mesure), au delà de laquelle toute modification de l'interligne articulaire pourrait être attribuée aux changements dus à la maladie.

CHAPITRE IV

REPRODUCTIBILITE

4-1)- DEFINITION ET INTERETS

La reproductibilité est la capacité d'une méthode de mesure à fournir des résultats semblables lors de mesures répétées dans des conditions identiques (*Lequesne 1989*) (*Leclercq 1990*) (*Bellamy 1993*) (*Dougados 1996*). Elle est synonyme de précision et est un reflet de l'erreur de mesure (*Streiner 1995*).

Il existe deux types de reproductibilité :

- La reproductibilité intra-observateur dans laquelle les mesures sont effectuées plusieurs fois sur les mêmes individus ou le même matériel par le même observateur.
- La reproductibilité inter-observateur dans laquelle les mesures sont effectuées par plusieurs observateurs sur les mêmes individus ou le même matériel.

La reproductibilité peut être transversale, lorsque qu'elle s'intéresse à la mesure à un temps donné, ou longitudinale, lorsqu'elle s'intéresse à la différence observée entre deux mesures effectuées après un intervalle de temps.

La reproductibilité est une condition préalable pour tout critère de jugement.

Une mesure ou un instrument doit être reproductible pour plusieurs raisons (*Wright 1992*). Une différence détectée par un instrument de mesure reproductible chez un patient peut être attribuée à un vrai changement de l'état clinique de ce patient. En pratique clinique, une imprécision (absence de reproductibilité) est source d'erreur, en particulier dans toutes les phases où la prise en charge du patient fait appel à la quantification d'un phénomène puis à une interprétation binaire fondée sur des valeurs seuils (existence ou absence de progression de l'arthrose ; succès ou échec d'un traitement).

Par ailleurs, la comparaison des performances de procédures médicales, dont on se demande s'ils donnent les « mêmes résultats », est une situation fréquente. Il s'agit parfois de remplacer une technique d'examen coûteuse ou pénible mais utilisée depuis des années par une technique moins chère et rapide. Une telle comparaison permet d'effectuer les choix de manière rationnelle.

En recherche, un instrument de mesure reproductible peut améliorer l'efficacité des essais cliniques. Ainsi, l'utilisation d'un instrument de mesure non reproductible peut conduire à ne pas mettre en évidence une différence réelle entre deux traitements comparés dans un essai clinique (risque β). Un instrument non reproductible diminue donc l'efficacité des essais cliniques. La solution à ce problème consiste habituellement à augmenter le nombre de sujets nécessaires à inclure dans l'essai (*Tugwell 1982*). D'un autre côté, dans les essais

multicentriques, il est essentiel de diminuer la variabilité entre les observateurs, sinon il existe un risque, non seulement de ne pas mettre en évidence des différences existant réellement, mais aussi d'obtenir des conclusions erronées.

4-2)- MESURES STATISTIQUES DE LA REPRODUCTIBILITE

4-2-1/ Variable continue

* Coefficients de corrélation

Les coefficients de corrélation sont parfois utilisés pour évaluer la reproductibilité. Les coefficients de corrélation le plus souvent utilisés sont les coefficients de corrélation de Pearson, Spearman et de Kendall (*Bellamy 1992*). Leurs valeurs varient entre -1 et $+1$. Un test statistique significatif est attendu puisque les examens sont répétés sur les mêmes individus.

Cependant, les coefficients de corrélation ne mesurent pas la concordance entre deux mesures mais, plutôt la force d'une relation linéaire entre ces deux mesures, c'est à dire la variation dans le même sens comme le poids et la taille, ou dans un sens opposé comme le taux d'hémoglobine et le volume sanguin (*Lee 1989*) (*Bland 1986*). Pour illustrer cet aspect, Bellamy cite l'exemple où deux examinateurs évalueraient une même variable (*Bellamy 1993*). L'un des examinateurs pourrait trouver des valeurs systématiquement supérieures, par exemple 5 fois supérieures, à celles trouvées par le deuxième examinateur. Dans ce cas, la corrélation pourrait être parfaite ($r=+1$), alors qu'il apparaît de cet exemple que les deux examinateurs sont en désaccord.

Par ailleurs, la corrélation dépend du spectre des valeurs dans l'échantillon étudié. Plus le spectre est large, plus grand sera le coefficient de corrélation. Hors, il est souvent recommandé d'évaluer la reproductibilité d'une mesure sur l'échantillon le plus représentatif de la population considérée, donc de considérer le spectre de sujets le plus large (*Wright 1992*). De ce fait, on obtient le plus souvent une corrélation forte (*Bland 1986*).

Enfin, l'utilisation de plus de deux examinateurs conduit à les comparer par paires et donc de calculer plusieurs coefficients de corrélation. En effet, il n'existe pas de statistique de corrélation validée permettant de combiner ces différents coefficients. L'interprétation globale de ces différents coefficients peut être difficile.

* Comparaison de moyennes

La comparaison de moyennes est inadaptée pour évaluer la reproductibilité. Des valeurs moyennes statistiquement égales, obtenues par deux examinateurs sur les mêmes sujets, n'impliquent pas une bonne reproductibilité entre ces deux examinateurs. La comparaison de moyennes est plutôt pertinente pour évaluer un biais possible, c'est à dire, si l'un des

examineurs tend à obtenir des valeurs systématiquement différentes de celles de l'autre examinateur (Lee 1989).

** Coefficients de variation*

Le coefficient de variation (CV) peut être calculé selon la formule suivante :

$CV = (SD/M) * 100$ où SD est la déviation standard et M la moyenne des mesures (Bellamy 1993).

Le coefficient de variation est fréquemment encore utilisé pour évaluer la reproductibilité alors qu'il est critiqué pour cette question (Feinstein 1985) (Bellamy 1993) (Tammenagi 1996). En effet, deux méthodes différentes peuvent avoir des coefficients de variation identiques alors qu'elles diffèrent pour leurs moyennes et déviations standards (Bellamy 1993).

Pour Feinstein, le coefficient de variation est un indice de compacité des données plutôt qu'un test de reproductibilité (Feinstein 1985).

Pour Tammenagi, l'utilisation de ce coefficient est responsable d'une perte d'information, puisqu'il ne prend pas en compte le fait que les données proviennent des mêmes sujets (données appariées) (Tammenagi 1996).

Enfin, Bellamy pense que le coefficient de variation ne doit plus être considéré comme mesure de la reproductibilité (Bellamy 1993).

** Coefficient de corrélation intra-classe*

Le coefficient de corrélation intra-classe a été proposé par Ebel puis par Bartko pour mesurer la reproductibilité des variables quantitatives (Ebel 1951) (Bartko 1966).

La définition du coefficient de corrélation intra-classe fait appel à l'idée de décomposition de la variance (Lapointe 1995). Dans le cas où 5 sujets sont examinés indépendamment par deux examinateurs à un moment donné, la variation totale entre les 10 examens est composée de la variation entre les examinateurs, de la variation entre les sujets (quelque soit l'examineur, les scores varient entre les sujets) et la variation dite résiduelle (non expliquée par le reste de la variation).

La variation inter-sujet correspond à une sorte d'effet sujet, dont le calcul nécessite au préalable que soient neutralisées les différences entre examinateurs. Pour cela, il est habituel

de considérer la moyenne des scores par sujet. L'effet sujet peut être alors mesuré par la différence entre le score moyen par sujet et le score moyen général.

Ainsi, le score attribué à un sujet par un examinateur est égal à la somme du score moyen général auquel s'ajoute l'effet sujet et une part résiduelle, soit :

$$\text{Score} = \text{score moyen général} + \text{effet sujet} + \text{résidu.}$$

Le résidu est nul quand l'accord entre les examinateurs est parfait. Le degré d'accord entre les examinateurs est donc un rapport entre l'effet sujet et le résidu. Si l'effet sujet explique la totalité de la variation, l'accord entre les examinateurs est parfait. Par contre, l'accord est mauvais quand la variation totale n'est en rien expliquée par l'effet sujet. Cette relation entre la variation inter-sujet et la variation totale (variation inter-sujet+variation résiduelle) permet d'envisager le rapport entre la variation inter-sujet et la variation totale comme mesure d'accord.

Ainsi le coefficient de corrélation intra-classe est mesuré par :

$\rho = \text{variation inter-sujet} / (\text{variation inter-sujet} + \text{variation résiduelle})$, soit de façon plus formelle :

$$\rho = \sigma_s^2 / (\sigma_s^2 + \sigma_e^2), \text{ où } \sigma_s^2 \text{ est la variance entre les sujets et } \sigma_e^2, \text{ la variance résiduelle.}$$

Armitage et Berry définissent alors le coefficient de corrélation intra-classe comme la corrélation entre chacune des paires individuelles de mesures par des méthodes choisies aléatoirement (*Armitage 1994*).

Ses valeurs varient entre 0 (absence totale d'accord entre les examinateurs, la variation totale est due aux seules erreurs aléatoires) et +1 (accord parfait entre les examinateurs, la variation totale est entièrement expliquée par la variation inter-sujet) (*Fermanian 1984b*) (*Lapointe 1995*).

En pratique, l'estimation du coefficient de corrélation intra-classe se fait à partir d'un tableau d'analyse de variance (modèle à effets aléatoires à un facteur) selon la formule suivante :

$r = (\text{CME} - \text{CMR}) / (\text{CME} + \text{CMR})$, où CME est le carré moyen correspondant au facteur sujet et CMR, le carré moyen résiduel (*Shrout 1979*) (*Fermanian 1994b*) (*Lapointe 1995*).

En cas de biais entre les mesures dans l'évaluation de la reproductibilité inter-lecteur, un modèle à deux facteurs (facteur sujet et facteur lecteur) doit lui être préféré (*Bartko 1966*) (*Shrout 1979*).

- Avantages et les inconvénients du coefficient de corrélation intra-classe :
 - Le coefficient de corrélation intra-classe a l'avantage de ne dépendre ni du niveau des mesures ni de l'ordre des répétitions (*Papoz 1994*).
 - Il est dépourvu d'unité et, variant de 0 (absence d'accord) à +1 (accord parfait), facile à interpréter (*Lee 1992*). Ainsi, Burdock et collaborateurs considèrent qu'une méthode dont la borne inférieure de l'intervalle de confiance unilatéral du coefficient de corrélation intra-classe est supérieure ou égale à 0,75 a une bonne reproductibilité, quel que soit la variable mesurée. (*Burdock 1963*).
 - Cependant, le coefficient de corrélation intra-classe combine l'information concernant une différence systématique éventuelle entre les méthodes comparées (biais) et la variation aléatoire de la mesure (erreur de mesure). Ce qui rend difficile son utilisation pour comparer des études (*Armitage 1994*).
 - Comme il est dépourvu d'unité, son utilisation est subjective quand il s'agit de comparer la reproductibilité de méthodes mesurant des variables différentes (*Lee 1992*).
 - Il dépend de la variabilité inter-sujet (*Bland 1986*). Ainsi, ce coefficient n'est généralisable qu'à la population dont provient l'échantillon étudié, si cet échantillon est représentatif. Le coefficient de corrélation intra-classe obtenu dans un échantillon ne doit pas être comparé à celui obtenu dans un sous-groupe de cet échantillon parce que leurs variances inter-sujet ne sont pas comparables (exemple, ensemble des mesures de la pression artérielle et mesures de la pression artérielle chez les sujets hypertendus ; ensemble des mesures de la pression artérielle systolique plus hétérogène que l'ensemble des mesures de la pression artérielle diastolique) (*Lee 1992*). En fait, le coefficient de corrélation intra-classe est influencé par la sélection des sujets. Si le spectre est large, le coefficient de corrélation intra-classe aura tendance à être grand, tandis qu'il sera plus faible dans un groupe de sujets plus homogène (*Armitage 1994*) (*Müller 1994*).
- Interprétation du coefficient de corrélation intra-classe : Burdock considère qu'une méthode a une bonne reproductibilité quand la borne inférieure de l'intervalle de confiance unilatéral de son coefficient de corrélation intra-classe est supérieure à

0,75 (*Burdock 1963*). Ce choix est arbitraire et le coefficient de corrélation intra-classe est d'autant plus grand que la variation entre les sujets est forte (*Bland 1990*) (*Armitage 1994*) (*Müller 1994*).

* Méthode de Bland et Altman

Cette méthode a été proposée dès 1986 par ces deux auteurs pour comparer des examens médicaux quand la variable est continue (*Bland 1986*) (*Bland 1990*) (*Bland 1995*) (*Bland 1999*).

Il s'agit d'une méthode à la fois statistique et graphique permettant de comparer aussi bien des mesures réalisées par deux examinateurs ou par deux examens chez un même sujet (reproductibilité inter-observateur) que des mesures répétées par un observateur ou un examen unique chez un même sujet (reproductibilité intra-observateur).

L'approche graphique consiste à représenter les différences (A-B) des résultats fournis par les paires de mesures A et B en ordonnée en fonction de la moyenne de ces mesures $(A+B/2)$ en abscisse. Ce graphique permet de :

- visualiser les erreurs de mesure en fonction de la valeur des résultats et ainsi de repérer les zones où ces erreurs sont élevées,
- donner un ordre de grandeur de l'erreur de mesure facilement interprétable par le clinicien.

Cette méthode est donc plus informative que la seule valeur du coefficient de corrélation intra-classe.

Elle permet d'estimer la moyenne des différences de paires de mesures (d) et l'écart-type (s) de ces différences.

Il existe un biais significatif si la moyenne des différences (d) est statistiquement différente de 0. Ce biais peut être ajusté en soustrayant (d) de la valeur obtenue par l'observateur qui surévalue.

L'erreur de mesure est représentée par l'écart-type (s) des différences entre les mesures.

On peut ensuite estimer l'intervalle dans lequel la plupart des désaccords entre les observateurs surviennent. Par exemple, 95 % des désaccords (différences) surviennent dans un intervalle compris entre $d-2S$ et $d+2S$, quand la distribution des valeurs est normale. Il s'agit des limites d'agrément. Quand la distribution n'est pas normale ou que l'échantillon est petit ($n < 30$), on utilise le nombre approché de degrés de liberté et l'intervalle est alors défini par $d \pm t_{n-1}S$ (où t_{n-1} correspond à la valeur de t pour $n-1$ degrés de liberté). Les limites d'agrément, tout comme (d), étant seulement des estimations des valeurs de la population

générale, on peut estimer leurs erreurs standards et leurs intervalles de confiance pour évaluer leur précision. Dans l'hypothèse de distribution normale, l'erreur standard de (d) est $\bullet S^2/n$, n étant la taille de l'échantillon, et l'erreur standard de $d-2S$ ou $d+2S$ est approximativement $\bullet(3S^2/n)$. On peut alors estimer les intervalles de confiance à 95 % des bornes supérieure et inférieure des limites d'agrément.

Dans l'exemple rapporté par Brennan et Silman où deux rhumatologues mesurent la force de préhension de 20 patients, les différences entre les mesures ne varient pas en fonction de la valeur des résultats $((A+B)/2)$ (Brennan 1992). Ce point est important car la méthode de Bland et Altman suppose que le niveau d'erreur est constant (Bland 1986). Par contre, dans cet exemple il existe un biais significatif, l'un des rhumatologues obtenant systématiquement des valeurs supérieures à celles du deuxième. Ce biais peut être la cause de la variation entre les deux examinateurs. En cas de biais, le calcul des intervalles de confiance et des limites d'agrément est inadapté. Cependant, on peut résoudre cette difficulté en procédant à une transformation logarithmique des données et cette relation disparaît.

La méthode proposée par Bland et Altman peut être également utilisée dans le cas où le nombre de répétition des mesures est supérieur à 2. Le graphique visualise alors l'écart-type des différentes mesures d'un sujet donné en fonction de leur moyenne.

Cependant, la méthode de Bland et Altman est critiquée car les résultats sont exprimés dans une unité de mesure rendant difficiles les comparaisons de reproductibilité de méthodes ayant des unités de mesure différentes (Lee 1992).

Enfin, l'interprétation des limites d'agrément est arbitraire puisqu'elle repose sur le jugement des utilisateurs des méthodes évaluées.

4-2-2/ Variable qualitative

*** Le test du Chi²**

Ce test est parfois utilisé pour mesurer la reproductibilité inter-observateur. Cependant, ce test mesure plutôt une liaison entre les modes de jugement et non leur accord (Fermanian 1984a).

*** Le coefficient Kappa**

Ce coefficient a été initialement développé par Cohen pour évaluer l'accord entre juges pour des variables dichotomiques ou ordinales (Cohen 1960). Il mesure le degré d'accord au-delà de celui lié au hasard (Cohen 1960) (Bellamy 1993) (Strenier 1995).

En épidémiologie, il s'agit du coefficient le plus souvent utilisé lorsque la variable à évaluer est qualitative. Dans le cas le plus simple, la variable est à deux niveaux (présence ou absence ; positif ou négatif ; normal ou anormal) : la mesure d'accord la plus simple est la proportion globale observée, p_o , de jugements concordants. Cependant, cette mesure est fortement influencée par la distribution des sujets positifs (normaux) et négatifs (anormaux). En cas de prépondérance de l'une des classes, une concordance élevée uniquement due au hasard peut être observée (*Steirner 1995*). Il est en effet possible d'imaginer un certain accord entre deux juges indépendants utilisant des jugements différents pour classer des sujets. On peut alors considérer que la valeur p_o , en comprenant les accords attribuables au hasard, exagère la proportion d'accords véritables.

On peut calculer la proportion globale d'accords due au hasard, p_e , à partir des distributions marginales d'un tableau de contingence (Tableau I).

Tableau I :

TABLEAU DE CONTINGENCE 2 x 2 RAPPORTANT LE CLASSEMENT DE N SUJETS
SELON 2 CATEGORIES :

		Juge 2		
Juge 1	Présent	Absent	TOTAL	
Présent	a	b	a + b	
Absent	c	d	c + d	
TOTAL	a + c	b + d	n	

p_o = somme des proportions observées de sujets classés « présent » ou « absent » par les deux juges soit $(a+d)/n$

p_e = somme des proportions de sujets classés « présent » ou « absent » due au hasard soit :

$$((a+b) \times (a+c))/n + ((c+d) \times (b+d))/n$$

Ainsi, il est possible de prendre en compte p_c en corrigeant p_o de l'effet attribuable au hasard de la façon suivante :

$p_o - p_c$ qui est une mesure possible d'accords véritables.

Plusieurs situations peuvent être observées :

1. $p_o > p_c$: les deux juges font mieux que le hasard, il existe un certain accord véritable.
2. $p_o = p_c$: les deux juges ne font pas mieux que le hasard, il n'y a pas d'accord véritable.
3. $p_o < p_c$: les deux juges font pire que le hasard.

La valeur de p_o varie entre 0 et 1, celle de $p_o - p_c$ est comprise entre $-p_c$ et $1 - p_c$.

Si $p_o - p_c > 0$, il existe un certain accord véritable. Pour savoir s'il s'agit d'un bon accord, on compare la valeur obtenue de $p_o - p_c$ à la valeur maximale dans toute condition, c'est-à-dire à $1 - p_c$.

Ainsi, le coefficient Kappa est défini par :

$$K = (p_o - p_c) / (1 - p_c)$$

Dans l'exemple hypothétique donné par Brennan et Silman, deux rhumatologues classent des radiographies des mains de 100 patients pour la présence ou l'absence d'érosions (Brennan 1992). La proportion globale observée de jugements concordants est de 70/100 soit 0,70, alors que la proportion globale due au hasard p_c (proportion attendue) obtenue à partir des valeurs marginales est de 0,54. La valeur de Kappa est alors de $(0,70 - 0,54) / (1 - 0,54)$, soit environ 0,35.

Des valeurs de Kappa égale à 0 indiquent une concordance due uniquement au hasard et +1, une concordance parfaite. Des valeurs négatives sont possibles quand la concordance observée semble inférieure au hasard (Kramer 1981) (Fleiss 1982) (Brennan 1992). Cependant, elles ont une probabilité faible de survenir en pratique (Armitage 1994).

Des règles d'interprétation du coefficient Kappa ont été proposées par plusieurs auteurs (tableau II) (Landis 1977) (Fleiss 1982). Elles permettent d'apprécier la force de la concordance à partir de la valeur du coefficient Kappa.

Tableau II :

DEGRE DE CONCORDANCE SELON LA VALEUR DU COEFFICIENT KAPPA SELON
LANDIS ET KOCH (*Landis 1977*) et selon Fleiss (*Fleiss 1982*) :

Valeurs du Kappa	Landis et Koch	Fleiss
0,81-1,00	Presque parfait	Excellent
0,76-0,80	Bon	Excellent
0,61-0,75	Bon	
0,41-0,60	Modéré	Modéré à bon
0,21-0,40	Médiocre	
0-0,20	Mauvais	Mauvais
< 0	Très mauvais	

* Le coefficient Kappa pondéré

Dans le cas où l'accord est évalué pour plus de 2 juges ou pour des variables qualitatives à plus de 2 classes, les possibilités d'erreurs et de désaccords augmentent avec le nombre de classes (*Brennan 1992*). Il est donc possible d'utiliser le coefficient Kappa pondéré permettant d'ajuster sur les différents niveaux de désaccords (*Cohen 1968*).

En effet, certaines discordances entre juges sont plus graves que d'autres. Cohen propose d'attribuer à chacune des cases du tableau de contingence des accords et des désaccords, un poids reflétant l'importance que l'on accorde à l'accord (ou au désaccord) reflété par la case (*Cohen 1968*). Le choix des poids se fait habituellement de manière telle qu'ils varient entre 0 et 1 (*Brennan 1992*).

Plusieurs systèmes de pondération ont été proposés (*Fernamian 1984*).

En pratique, quand il y a accord entre les juges, le poids W_i est égale à 1. En cas de k classes, le désaccord maximal est de $k-1$ classes et le poids prend la valeur 0. Le système de poids le plus fréquemment utilisé pour les valeurs intermédiaires est :

$$W_i = 1 - (i/K - 1), \text{ où } i \text{ est le nombre de classes discordantes.}$$

Le choix des pondérations influence les résultats (*Brennan 1992*). Il est donc difficile de comparer les résultats obtenus dans des études similaires mais avec des systèmes de pondération différents (*Maclure 1987*) (*Steiner 1995*).

Le coefficient Kappa pondéré KW se définit par :

$(P_o - P_c)/(1 - P_c)$ où P_o et P_c sont estimés en utilisant les pondérations (*Cohen 1968*) (*Brennan 1992*) (*Fernamian 1984a*).

Il a les mêmes propriétés que le coefficient Kappa et varie entre 0 et +1.

Le coefficient Kappa pondéré et le coefficient de corrélation intra-classe sont liés. Leurs valeurs sont égales si on utilise un système de poids $W_i = 1 - (i^2 / (K-1)^2)$ pour le Kappa pondéré (*Fleiss 1973*) (*Armitage 1994*).

* Intérêt et limites du coefficient Kappa

Ces coefficients permettent une évaluation honnête de la reproductibilité inter-juge aussi bien dans des disciplines apparemment objectives telles que la radiologie, que dans des disciplines faisant appel à plus de subjectivité dans le jugement. Ils permettent de mettre en évidence une discordance imprévue dont on recherchera la cause ou d'objectiver et de mesurer une discordance soupçonnée.

Ils peuvent aussi révéler l'existence d'un biais. Une discussion entre juges peut alors permettre de l'objectiver. La mesure d'une discordance non systématique permet de mettre en œuvre une stratégie pour améliorer la reproductibilité.

Cependant, deux paradoxes du coefficient Kappa ont été décrits (*Feinstein 1990*) (*Cicchetti 1990*). Premièrement, pour une même valeur P_o élevée, le coefficient Kappa peut être élevé si la valeur de P_c est basse, ou faible si la valeur de P_c est élevée. Deuxièmement, K sera plus élevé en cas de déséquilibre asymétrique ou imparfait des valeurs marginales.

Enfin, Kappa dépend de la prévalence du paramètre étudié dans l'échantillon et du niveau de discordance entre juges (*Armitage 1994*). Ainsi, la comparaison des coefficients Kappa entre études est difficile à interpréter. Un indicateur, le PABAK (Prévalence Adjusted Bias-adjusted Kappa), a été proposé pour prendre en compte la prévalence et le niveau de discordance (*Byrt 1993*).

CHAPITRE V

VARIABILITE DE LA MESURE DE L'INTERLIGNE ARTICULAIRE

5-1)- ETUDE DE LA CONCORDANCE ENTRE LES MESURES DE L'INTERLIGNE ARTICULAIRE EFFECTUEES SUR DES RADIOGRAPHIES REALISEES EN POSITIONS COUCHEE ET DEBOUT

5-1-1/ Etat des connaissances et objectifs de la recherche

Dans les études épidémiologiques évaluant la coxarthrose, les radiographies standards sont réalisées sur des patients en position debout (*Dougados 1996a*) ou couchée (*Altman 1987*) (*Ledingham 1992*) (*Ledingham 1993*). L'intérêt de la mise en charge du patient pendant la radiographie du bassin reste discuté (*Resnick 1988*) (*Dieppe 1994*) (*Dieppe 1995*) (*Altman 1996*).

Par ailleurs, les recommandations récentes proposent d'évaluer la coxarthrose sur des radiographies standards réalisées en position debout (*Altman 1996*), couchée (*Dieppe 1994*), ou dans les deux positions (*Dieppe 1995a*).

Les études évaluant l'effet de la mise en charge sur la hauteur de l'interligne articulaire de hanche sont peu nombreuses et ont des résultats contradictoires.

Conrozier et collaborateurs ont comparé les mesures de l'interligne articulaire (surface, épaisseur moyenne et hauteur au point le plus pincé) sur des radiographies réalisées en position debout et couchée (*Conrozier 1997*). Cette étude comportait deux parties : radiographies du bassin (46 hanches arthrosiques et 11 hanches saines) et radiographies de hanche (28 hanches arthrosiques). Pour l'ensemble des patients, il n'y avait aucune différence statistique entre les mesures obtenues sur les radiographies réalisées en position debout et couchée. Cependant, dans le sous-groupe des hanches ayant un interligne inférieur ou égal à 2,5 mm, ces auteurs ont montré que la hauteur de l'interligne au point le plus pincé était significativement plus petite sur les radiographies réalisées en position debout par rapport à celles réalisées en position couchée (de 6,7% sur les radiographies du bassin et de 11,5% sur les radiographies de hanche).

Hansson et collaborateurs ont, quant à eux, étudié 151 hanches (*Hansson 1993*). La hauteur moyenne de l'interligne articulaire était significativement plus petite sur les radiographies réalisées en position debout ($3,9 \pm 1,7$ mm) que sur les radiographies réalisées en position couchée ($4,2 \pm 1,7$ mm).

Evison et collaborateurs ont étudié 13 hanches arthrosiques et 25 hanches saines (*Evison 1987*). Cependant, seules 2 hanches arthrosiques et 1 hanche saine avaient un interligne plus

petit sur les radiographies réalisées en position debout que sur les radiographies réalisées en position couchée.

Plus récemment, Pessis et collaborateurs ont comparé les mesures de l'interligne articulaire sur ces deux types de radiographies réalisées chez 25 patients (23 hanches arthrosiques et 27 hanches saines) (Pessis 1999). Ils n'ont pas montré de différence significative entre les mesures de l'interligne articulaire, ni pour l'ensemble des hanches ni dans les sous-groupes de hanches saines et arthrosiques.

Enfin, dans leur étude comparant l'interligne des deux hanches de 20 patients, Teshima et collaborateurs ont montré que la surface de l'interligne articulaire était plus petite d'au moins 10 % à 30 % du côté en décharge (depuis 5 à 92 mois) que du côté resté en charge (Teshima 1992).

Par ailleurs, dans certaines de ces études, pendant la réalisation des radiographies, la rotation des membres inférieurs des patients pouvait varier avec la position couchée ou debout (Hansson 1993), et les mesures de l'interligne articulaire n'étaient pas optimales (compas) (Evison 1987). Enfin, les méthodes statistiques utilisées dans la plupart de ces études étaient inadéquates au problème de la reproductibilité.

Ainsi, les études épidémiologiques et les essais cliniques utilisant des radiographies réalisées en position debout ne peuvent être comparés à ceux utilisant des radiographies réalisées en position couchée, si les mesures obtenues sur ces deux types de radiographies ne sont pas concordantes.

Nous avons donc tenté de comparer les mesures de la hauteur de l'interligne articulaire sur les radiographies standards du bassin réalisées en positions debout et couchée, chez des patients souffrant ou non de coxarthrose.

Nous avons utilisé la méthode proposée par Bland et Altman pour analyser les résultats comparant deux examens mesurant le même phénomène.

Nous avons montré qu'il n'existait pas de biais (différence systématique) entre les mesures de la hauteur de l'interligne articulaire, obtenues sur des radiographies réalisées en position debout et couchée, ni pour l'ensemble des hanches ni dans les sous-groupes de hanches saines et arthrosiques.

Cette étude nous a donc permis de proposer des limites d'agrément de 0,64 mm entre les interlignes mesurées sur les radiographies réalisées dans ces deux positions.

5-1-2/ Article 1: Osteoarthritis of the hip: agreement between joint space width measurements on standing and supine conventional radiographs.

G-R Auleley¹, B Rousselin², X Ayrat¹, R Edouard-Noel¹, M Dougados¹, P Ravaud¹

1- Clinique de Rhumatologie, Hôpital Cochin, Paris, France.

2- Service de Radiologie, Hôpital Ambroise Paré, Paris, France.

Ann Rheum Dis 1998; 57: 519-23.

EXTENDED REPORTS

Osteoarthritis of the hip: agreement between joint space width measurements on standing and supine conventional radiographs

Guy-Robert Auleley, Benoît Rousselin, Xavier Ayrat, Rachel Edouard-Noel, Maxime Dougados, Philippe Ravaud

Abstract

Objective—To assess the effect of standing position on joint space width (JSW) measurements of the hips with and without osteoarthritis (OA) on pelvic radiographs.

Methods—Adult patients aged 18 or more had pelvic anteroposterior conventional radiographs standing and supine performed by a single radiologist in the same radiology unit according to standardised guidelines. JSW measurements in mm were made by a single reader blind to patients' identity and type of view, using a 0.1 mm graduated magnifying glass directly laid over the radiograph, at the narrowest point for OA hips or at the vertical joint space for non-OA hips. Agreement of JSW between both views was assessed using the Bland and Altman graphical analysis.

Results—JSW was greater on standing than supine radiographs, for example, 7.1% for OA hips. Mean (SD) differences and limits of agreement (mm) between both views were 0.08 (0.27) and -0.46 to 0.62 for the 70 non-OA hips, 0.02 (0.31) and -0.60 to 0.64 for the 46 OA hips. Corresponding 95% confidence intervals of mean differences were 0.02 , -0.14 mm and -0.07 , -0.11 mm.

Conclusions—Measurements of JSW of the hip on pelvic standing and supine radiographs are concordant. Changes less than or equal to 0.64 mm between the two views are similar or inferior to radiological progression of OA.

(Ann Rheum Dis 1998;57:519-523)

standing position.¹⁰ Examination of the hip by pelvic radiograph while the patient is bearing weight has generally not been considered useful except for evaluation of the postoperative hip.¹¹ In recent guidelines for evaluating OA, it has been recommended that hip radiographs be performed with the patient standing,⁶ or supine,¹² or either.³ In OA of the knee JSW is narrower on weightbearing than on non-weightbearing radiographs.^{7 13 14} Furthermore, JSW of the medial compartment of the osteoarthritic knee is accurately measured on weightbearing radiographs.¹⁵ Studies comparing JSW of the hip measured on weightbearing and non-weightbearing radiographs have given confusing results.¹⁶⁻¹⁹ Conrozier *et al*¹⁶ found that maximum joint space narrowing of the hip significantly decreased on weightbearing views compared with supine views only when considering OA hips with JSW less than or equal to 2.5 mm. Evison *et al*¹⁷ found that compared with non-weightbearing views, hip JSW decreased in two of 13 OA hips and increased or decreased in three of 25 non-OA hips on weightbearing views. For these authors, the latter view may be useful in selected groups of patients. Hansson *et al*¹⁸ have reported narrowing of JSW on weightbearing radiographs in 12 of 117 hips considered normal on non-weightbearing views. Of the 34 hips considered as narrowed on non-weightbearing radiographs only one became normal on weightbearing views in their study. Therefore, they did not recommend supine radiographs for assessing coxarthrosis. Other authors have shown radiographic narrowing of the joint space of 10% to 30% or more in hips weight free for 5 to 92 months as compared with weightbearing hips.¹⁹ In routine practice, hip OA is assessed on radiographs performed in supine or standing conditions. Consequently, the possibility of comparing epidemiological studies and clinical trials using radiographs performed with either type of patient positioning is limited if radiographs obtained with both patient positionings for measuring JSW of hip OA are not interchangeable.

The aim of this study was to assess the effect of the standing position on JSW measurements of the hip on pelvic radiographs.

Assessment of articular loss has been shown to be important for evaluating the progression of osteoarthritis (OA) in epidemiological studies and clinical trials.¹ A measure of articular cartilage is considered the primary outcome variable of OA for structure modifying drugs. Several guidelines recommend that the measure of joint space width (JSW) determined on radiography should be used as a proxy for joint space narrowing.²⁻⁶ To study OA hips, radiographs are obtained with patients in supine⁷⁻⁹ or

Faculté de Médecine
Cochin Port-Royal and
Clinique de
Rhumatologie, Hôpital
Cochin, AP-HP, Paris,
France
G-R Auleley
X Ayrat
R Edouard-Noel
M Dougados
P Ravaud

Service de Radiologie,
Hôpital Ambroise
Paré, AP-HP, Paris,
France
B Rousselin

Correspondence to:
Dr P Ravaud, Clinique de
Rhumatologie, Hôpital
Cochin, 27 rue du faubourg
Saint Jacques, 75014 Paris,
France.

Accepted for publication
16 July 1998

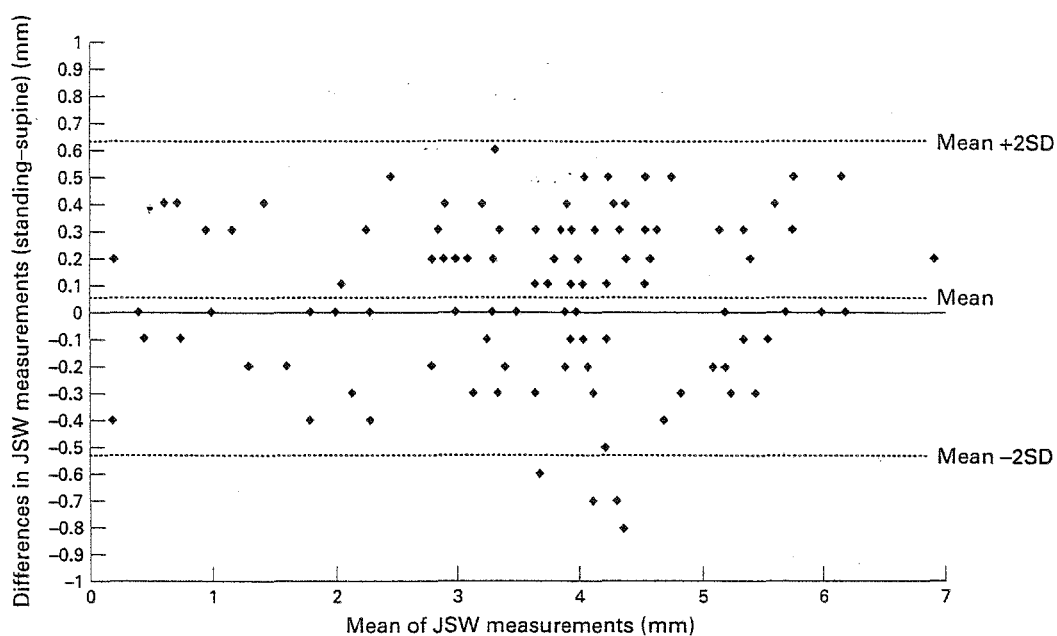


Figure 1 Differences in joint space width (JSW) (in mm) between standing and supine radiographs against the mean for all hips. Limits of agreement correspond to mean \pm 2SD.

Methods

PATIENTS

Consecutive patients of both sexes referred to a radiology unit for radiographic evaluation of the pelvis were eligible for the study. Other inclusion criteria were: age equal to or greater than 18 years, referral for chronic back pain, sciatica, bone neoplasia, myeloma or suspicion of hip OA. Patients were excluded if they had secondary hip OA defined by the presence or past history of hip fracture, inflammatory rheumatic disease, osteonecrosis, Paget's disease, or infectious disease.

RADIOGRAPHS

A single radiologist performed all radiographs in the same radiology unit according to a standardised procedure. For each patient, two conventional pelvic radiographs were made. The first film was performed in standing conditions. Anteroposterior radiographs were taken with a source to film distance of 110 cm. The patient's feet were internally rotated with the toes at $15 \pm 5^\circ$ with the help of fluoroscopy to ascertain that the x ray beam was centred on the superior aspect of the pubic symphysis. The x ray beam was horizontal, perpendicular to the table. Patients were standing on both legs and were asked to distribute their weight equally on both feet.

After patients gave informed consent, the second film was performed in supine conditions using the same radiological procedure (incidence, source to film distance, rotation of patient's feet) as for standing films except that the x ray beam was vertical.

The radiologist (BR) classified radiographs as with hip OA (definite osteophytes and joint space narrowing)²⁰ or not (absence of osteophytes and joint space narrowing).

RADIOGRAPHIC ASSESSMENT

All the radiographs were collected and the patients' identity and the type of view were masked with an adhesive tape by one investigator (REN). The patients' identification was replaced by a random code number. Radiographs were assessed blind to the view performed by one reader (GRA). Before masking, the two radiographs of a single patient were placed side by side on a light box and the landmarks for measurements were drawn by the reader. The landmarks consisted of two points, one on the distal margin of the condylar cortex for the femoral surface and the other on the margin of the bright radiodense band of the subchondral cortex in the floor of the articular fossa for the acetabulum. This interbone distance was measured in millimetres using a 0.1 mm graduated magnifying glass directly laid over the radiograph at the narrowest point for OA hips and at the vertical joint space for non-OA hips. In addition, for OA hips joint space narrowing was graded 0–3 using a radiographic atlas²¹ and overall severity of OA was graded using the Kellgren and Lawrence criteria.²²

In radiographs classified by BR as without hip OA, both hips were examined except for two radiographs with unilateral prosthetic hip related to femoral neck fracture. In radiographs classified by BR as with hip OA, both hips were

Table 1 Means (standard deviations) and (range) (mm) of joint space width in 46 hips with osteoarthritis and 70 hips without osteoarthritis on pelvic standing and supine radiographs during two reading sessions

	Standing without osteoarthritis (n=70)	Supine without osteoarthritis (n=70)	Standing with osteoarthritis (n=46)	Supine with osteoarthritis (n=46)
First reading session	4.51 (0.90) (2.7–7.0)	4.43 (0.90) (2.7–7.0)	2.51 (1.33) (0–5.0)	2.50 (1.36) (0.1–4.7)
Second reading session	4.48 (0.92) (2.7–7.0)	4.40 (0.87) (2.5–6.5)	2.50 (1.35) (0.1–4.7)	2.50 (1.34) (0.2–4.7)

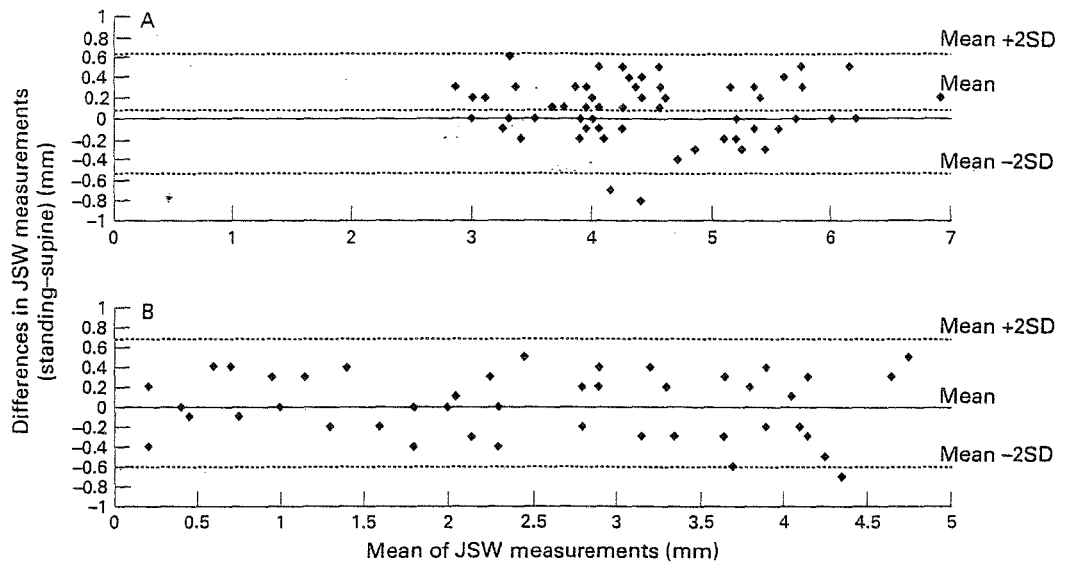


Figure 2 (A) Hips without osteoarthritis. Differences in joint space width (JSW) (in mm) between standing and supine radiographs against the mean for hips without osteoarthritis. Limits of agreement correspond to mean \pm 2SD. (B) Hips with osteoarthritis. Differences in joint space width (JSW) (in mm) between standing and supine radiographs against the mean for hips with osteoarthritis. Limits of agreement correspond to mean \pm 2SD.

examined except for 10 radiographs with unilateral prosthetic hip (two radiographs) or unilateral normal hip (eight radiographs). The reader (GRA) separately measured all JSW of OA and non-OA hips in one reading session.

A week later, the same reader, unaware of the results of the first reading session, measured all JSW of OA and non-OA hips. In this second reading session, the radiographs had their random code number and were evaluated in a different order.

STATISTICAL ANALYSIS

Means and standard deviations (SD) of JSW measurements in both views during each reading session, and their differences were calculated. JSW of hips measured on standing radiographs were also presented as a relative percentage of JSW measured on supine radiographs using the difference between JSW on supine and standing radiographs as the numerator and JSW on supine radiographs as the denominator.

The first step of analysis consisted in evaluating the agreement between JSW measurements on standing and supine radiographs. For this purpose, the magnitude of changes, including those related to the patient's position was assessed in all, non-OA, and OA hips. For all these groups, the graphical analysis described by Bland and Altman^{23, 24} was applied. This method focuses on differences between pairs of measurements of the same quantity against their corresponding mean. Graphs were plotted for each of these groups. Limits of agreement were estimated as $d_{\text{mean}} \pm 2\text{SD}$, where d_{mean} is the mean difference and SD, the standard deviation of the differences.

The second step consisted in evaluating the magnitude of changes that might result from measurement error. This magnitude was also estimated by using Bland and Altman graphical analysis on measurements repeated one week apart of OA and non-OA hips either in standing or supine radiographs.

Results

Sixty four patients (42 women and 22 men) were included. They had a mean age of 66 (range 38–92) years. A total of 116 hips, regardless of whether they were OA, were available for analysis. Means (SD) of JSW were 3.72 (1.45) mm and 3.66 (1.45) mm on standing and supine radiographs, respectively. The differences in measurements between the two views varied from -0.6 mm to 0.7 mm. Figure 1, plotting the difference between pairs of measurements against the corresponding mean of each hip, shows that there was no relation between the difference and the mean. On average, JSW was 4.1% (SD, 25.6%) greater on standing than on supine radiographs.

JSW was measured in 70 non-OA hips. Table 1 gives the means (SD) of JSW measurements. The differences in measurements between the two views varied from -0.6 mm to 0.7 mm and their variation did not depend on their mean (fig 2A). On average, JSW was 2.2% (SD, 6.3%) greater on standing than on supine radiographs.

Forty six OA hips were available for measurement of JSW. They were graded 2 (n=29), 3 (n=11) or 4 (n=6) in the Kellgren and Lawrence grading system and graded 1 (n=30), 2 (n=13) or 3 (n=3) in the joint space narrowing grading scale. Means (SD) of JSW measurements were identical between standing and supine radiographs (table 1). The differences in JSW between the two views varied from -0.6 mm to 0.7 mm. No relation existed between the difference and the mean (fig 2B). JSW was 7.1% (SD, 40%) on average greater on standing than

Table 2 Comparison of joint space width measurements on standing and supine radiographs of 46 hips with osteoarthritis

Standing	Supine		Total
	> 3.0 mm	≤ 3.0 mm	
> 3.0 mm	16	2	18
≤ 3.0 mm	1	27	28
Total	17	29	46

Table 3 Variability between views or between reading sessions of joint space width (mm) of hips with osteoarthritis and hips without osteoarthritis (OA) on radiographs performed in the same patients

	Number	Mean differences	Standard deviations	95% confidence intervals	Limits of agreement
Between views					
all	116	0.05	0.29	0.0, 0.10	-0.53 to 0.63
without OA	70	0.08	0.27	0.02, 0.14	-0.46 to 0.62
with OA	46	0.02	0.31	-0.07, 0.11	-0.60 to 0.64
Between reading sessions					
standing without OA	70	0.03	0.13	0.0, 0.06	-0.20 to 0.26
supine without OA	70	0.03	0.13	0.0, 0.06	-0.20 to 0.26
standing with OA	46	0.02	0.11	-0.01, 0.05	-0.20 to 0.24
supine with OA	46	-0.002	0.13	-0.04, 0.04	-0.26 to 0.26

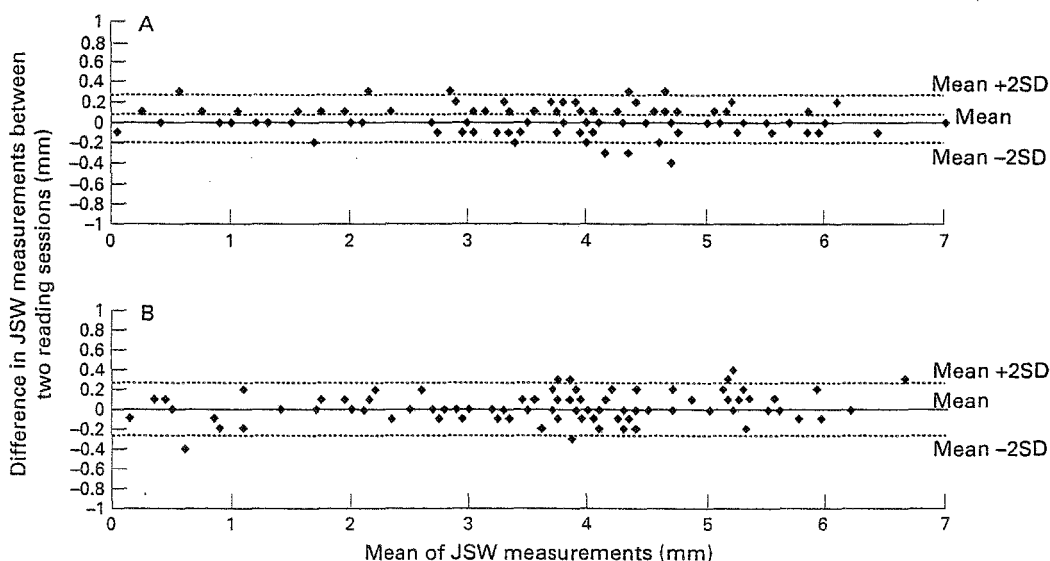


Figure 3 (A) Standing radiographs. Differences in joint space width (JSW) (in mm) for standing radiographs between two reading sessions against the mean for all hips. Limits of agreement correspond to mean \pm 2SD. (B) Supine radiographs. Differences in joint space width (JSW) (in mm) for supine radiographs between two reading sessions against the mean for all hips. Limits of agreement correspond to mean \pm 2SD.

supine radiographs. When considering the 23 hips with JSW less than or equal to 2.5 mm on the one hand and the 23 hips with JSW greater than 2.5 mm on the other hand, mean differences (SD) of JSW between the two views were 0.01 (0.26) mm and 0.02 (0.37) mm, respectively. When considering the limit for definite narrowing of joint space as less than or equal to 3 mm, one of the 17 hips with a normal joint space on the supine radiograph was reclassified on standing. Two of the 29 abnormal hips on the supine radiograph were reclassified normal on standing (table 2).

The largest magnitude of changes related to the measurement error was observed in non-OA hips (mean difference=0.03 mm) and the smallest in OA hips in supine radiographs (mean difference=-0.002 mm) (table 3). Comparison of JSW between two reading sessions showed wider limits of agreement for OA hips radiographed in supine conditions (-0.26 to 0.26 mm). No relation appeared between the measurement error and the estimated true value (fig 3).

Table 3 summarises also mean differences (SD), 95% confidence intervals, and limits of agreement. All confidence intervals overlapped zero.

Discussion

Measurements of JSW of hip with or without OA on conventional pelvic anteroposterior

radiographs performed in standing and supine conditions are concordant. Mean differences (SD) and limits of agreement between the two views were 0.08 (0.27) mm, -0.46 mm to 0.62 mm, and 0.02 (0.31) mm, -0.60 mm to 0.64 mm in non-OA and OA hips, respectively. These limits of agreement suggest that changes of less than 0.64 mm between JSW measurements on standing and supine radiographs are within the normal variability of measurements. Indeed, such variability is clinically relevant as it is similar to or less than a change in JSW proposed by others to define radiological progression of OA related to the course of the disease.^{10 25}

Mean differences in JSW were less than 0.1 mm in our study whereas those found statistically significant in other studies were 0.13 mm¹⁶ and 0.3 mm.¹³ The reader measured JSW blind to the patient's identity and to the type of view in our study. It is not stated whether others did the same when comparing the two views.¹⁶⁻¹⁹ When comparing methods of measurement, knowledge of the result by one method (lack of blind) may affect the result of the second measurement, resulting in a bias effect. The Bland and Altman graphical method permits this effect to be assessed along with the random effect.

As possible bias between repeated measurements can limit the amount of agreement,^{23 24 26} study of reproducibility of measurements on each view was relevant and allowed to

distinguish changes related to measurement error from changes related to patient's position.

We compared JSW measurements obtained on weightbearing and non-weightbearing radiographs using Bland and Altman graphical plots. This method is appropriate when comparing methods of clinical measurement.^{26 27} What we intended to assess was how well JSW measurements agree for individual people. Statistical methods like the test of significance for means are irrelevant to the question of agreement.^{28 29}

Our results differ from those found for the knee joint. Differences in biomechanics of the hip and the knee may explain such discrepancies. Deformations of the hip articular cartilage when just standing have not really been studied. However, the pressure exerted through the hip when standing with body weight equally distributed on both feet may be less than when walking.^{30 31} Furthermore, the pressure exerted on the hip may exceed half the body weight when subjects are lying, either prone or supine.³²

Our results differ from those found elsewhere, especially by Hansson *et al.*¹⁸ In our study, narrowing of the hip joint on standing radiographs was observed in one (2.9%) of the hips considered normal on supine radiographs. But narrowing of hip joint on supine radiographs was observed in two (11%) of the hips considered normal on standing radiographs. In contrast, Hansson *et al* found narrowing of the hip joint on standing radiographs in almost 10% of the hips considered normal on supine radiographs, whereas they found narrowing of the hip joint on supine radiographs in only 0.9% of the hips considered normal in standing radiographs. Several factors may explain such differences in results. Unlike us, Hansson *et al* did not use a blinding procedure during measurement of JSW. Moreover, in our study radiographs were performed with patients' feet internally rotated in standing and supine positions. Hansson *et al* performed radiographs with patients' feet internally rotated during supine radiographs and in neutral position during standing radiographs. It has been reported that rotation of the foot during the radiograph may influence JSW measurement.³³ Finally, our patients had pelvic radiographs while those included by Hansson *et al* had hip radiographs. The direction of the x ray beam may also contribute to differences in JSW measurements.^{6 33}

A limitation of this study was that we did not assess whether longitudinal radiographs performed in the same patient in standing and supine conditions are also concordant.

We have shown that standing and supine radiographs of the hip in the same patient allow identical measurements of JSW. Evaluating whether serial standing and supine radiographs of the hip allow identical JSW remains to be performed.

1 Jonsson K, Buckwalter K, Helvie M, Niklason L, Martel W. Precision of hyaline cartilage thickness measurements. *Acta Radiol* 1992;33:234-9.

- 2 Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. *Osteoarthritis Cartilage* 1995;3:73-7.
- 3 Group for the respect of ethics and excellence in science (GREES): Osteoarthritis section: Recommendations for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis* 1996;35:552-7.
- 4 Hochberg MC, Altman RD, Brandt KD, Moskowitz RW, for the Task force. Design and conduct of clinical trials in osteoarthritis: Preliminary recommendations from a task force of the Osteoarthritis Research Society. *J Rheumatol* 1997;24:792-4.
- 5 Lequesne M, Brandt K, Bellamy R, Moskowitz R, Menkes CJ, Pelletier J-P. Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol* 1994;21 (suppl 41):65-73.
- 6 Recommendations from a task force of the Osteoarthritis Research Society: Design and conduct of clinical trials in patients with osteoarthritis. *Osteoarthritis Cartilage* 1996;4:217-43.
- 7 Altman RD, Fries JF, Bloch DA, Cartens J, Cooke TD, Genant H, *et al*. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987;30:1214-25.
- 8 Ledingham J, Dawson S, Preston B, Milligan G, Doherty G. Radiographic patterns and associations of osteoarthritis of the hip. *Ann Rheum Dis* 1992;51:1111-16.
- 9 Ledingham J, Dawson S, Preston B, Milligan G, Doherty G. Radiographic progression of hospital referred osteoarthritis of the hip. *Ann Rheum Dis* 1993;52:263-7.
- 10 Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996;55:356-62.
- 11 Resnick D, Niwayama G. *Diagnosis of bone and joint disorders*. Vol 1. Philadelphia: WB Saunders, 1988:40.
- 12 Dieppe P, Altman RD, Buckwalter JA, Felson DT, Hascall V, Lohmander LS, *et al*. Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. In: Kuettner KE, Goldberg VM, eds. *Osteoarthritis disorders*. Rosemont, IL: American Academy of Orthopaedic Surgeons, 1994:481-96.
- 13 Ahlback S. Osteoarthritis of the knee: a radiographic investigation. *Acta Radiol* 1968;277 (suppl):7-72.
- 14 Leach RE, Gregg T, Siber FJ. Weight-bearing radiograph in osteoarthritis of the knee. *Radiology* 1970;97:265-8.
- 15 Buckland-Wright JC, Macfarlane DG, Lynch JA, Jasani MK, Bradshaw CR. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. *Ann Rheum Dis* 1995;54:263-8.
- 16 Conrozier T, Lequesne MG, Tron AM, Mathieu P, Berdah L, Vignon E. The effects of position on the radiographic joint space in osteoarthritis of the hip. *Osteoarthritis Cartilage* 1997;5:17-22.
- 17 Evison G, Reilly PA, Gray J, Calin A. Comparison of erect and supine radiographs of the hip. *Br J Rheumatol* 1987;26:393-4.
- 18 Hansson G, Jerre R, Sanders SM, Wallin J. Radiographic assessment of coxarthrosis following slipped capital femoral epiphysis: A 32-year follow-up study of 151 hips. *Acta Radiol* 1993;34:117-23.
- 19 Teshima R, Otsuka T, Yamamoto K. Effects of nonweight bearing on the hip. *Clin Orthop* 1992;279:149-56.
- 20 Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al*. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505-14.
- 21 Altman RD, Hochberg M, Murphy Jr WA, Wolphe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995;3(suppl A) 3-70.
- 22 Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16:494-501.
- 23 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10.
- 24 Bland JM, Altman DG. Comparing two methods of clinical measurement: A personal history. *Int J Epidemiol* 1995;24 (suppl 1):S7-14.
- 25 Dieppe PA, Cushnaghan J, Jasan MK, Mc Crea F, Watts I. A two-year placebo-controlled trial of non-steroidal anti-inflammatory therapy in osteoarthritis of the knee joint. *Br J Rheumatol* 1993;32:595-600.
- 26 Chinn S. The assessment of methods of measurements. *Stat Med* 190;9:351-62.
- 27 Atkinson G, Nevill A. Comment on the use of concordance correlation to assess the agreement between two variables. *Biometrics* 1997;51:775-7.
- 28 Bartko JJ. General methodology II: Measures of agreement: A single procedure. *Stat Med* 1994;13:737-45.
- 29 Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 1989;19:61-70.
- 30 Armstrong CG, Bahrani AS, Gardner DL. In vitro measurement of articular cartilage deformations in the intact human hip joint under load. *J Bone Joint Surg Am* 1979;61:744-55.
- 31 Denham RA. Hip mechanics. *J Bone Joint Surg Br* 1959;41: 550-7.
- 32 Bullough P, Goodfellow J, O'Connor J. The relationship between degenerative changes and load-bearing in the human hip. *J Bone Joint Surg Br* 1973;55:746-58.
- 33 Buckland-Wright JC. Quantitative of radiographic changes. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998.

5-2)- ETUDE DE L'INFLUENCE DE LA ROTATION DES MEMBRES INFÉRIEURS ET DU CENTRAGE DU RAYON INCIDENT

5-2-1/ Etat des connaissances et objectifs de la recherche

La dégradation du cartilage articulaire, mesurée par le pincement de l'interligne au point le plus pincé sur des radiographies standards, est le critère recommandé pour évaluer la progression de l'arthrose (Altman 1996). Cette progression est lente, de l'ordre de 0,1 à 0,6 mm par an (Mazières 1990b) (Kirwan 1992) (Lequesne 1995).

De nombreuses sources de variabilité peuvent compromettre la reproductibilité des mesures de l'interligne articulaire (technique d'acquisition des radiographies, position de l'articulation, instrument de mesure, site de mesure, lecteur) (Ravaud 1997).

Pendant la réalisation des radiographies explorant la hanche, la rotation des membres inférieurs des patients peut varier de 0° (Hansson 1993) à 25° (Conrozier 1997). Par ailleurs, elle n'est pas précisée dans certaines études (Croft 1990) (Ledingham 1993) (Dougados 1996a).

Les recommandations concernant l'évaluation de la coxarthrose sont contradictoires sur la position des membres inférieurs pendant la réalisation des radiographies. Resnick et Niwayama préconisent une rotation interne d'environ 15° (Resnick, 1988). Dieppe et collaborateurs recommandent que les orteils du patient soit en rotation interne de 10°±5° (Dieppe 1994). D'après les recommandations de l'OARS, la rotation des orteils doit être de 15°±5° (Altman 1996).

A notre connaissance, l'influence de la position des membres inférieurs du patient pendant la réalisation des radiographies standards n'a pas encore été évaluée dans la coxarthrose. Cependant, dans une étude de 9 cadavres, Buckland-Wright a observé que la hauteur de l'interligne articulaire était significativement inférieure sur des radiographies réalisées avec une rotation interne des membres inférieurs par rapport à celles réalisées à sans rotation (Buckland-Wright 1998).

D'autre part, les mesures de l'interligne articulaire sont parfois obtenues sur urographies intraveineuses (Croft 1990) ou des lavements barytés en double contraste (Ingvarsson 2000), utilisant un centrage du rayon incident très différent de celui des radiographies du bassin habituellement utilisées.

Cependant, l'effet du centrage du rayon sur les mesures de l'interligne articulaire est largement méconnu.

Nous avons donc étudié dans un premier temps l'effet de la modification de la position du patient (rotation interne des membres inférieurs) et de la technique radiologique (centrage du rayon incident).

Nous avons choisi d'étudier chacun des paramètres sur un groupe de patients différent :

- La rotation des membres inférieurs à 15° et 5°,
- Le centrage du rayon incident 2 cm au dessus du bord supérieur de la symphyse pubienne et sur l'ombilic.

Dans la seconde partie de l'étude, nous avons étudié l'influence de la standardisation de tout le processus depuis l'acquisition des images radiographiques jusqu'à la mesure de l'interligne articulaire.

La variabilité liée à la modification de la rotation des membres inférieurs ou du centrage du rayon incident d'une part et la reproductibilité de l'ensemble du processus standardisé d'autre part, ont été évaluées par la méthode proposée par Bland et Altman.

Nous avons montré qu'il existait un biais (différences significative) entre les mesures de la hauteur de l'interligne articulaire sur des radiographies réalisées avec un centrage du rayon sur l'ombilic et celles obtenues sur les radiographies réalisées avec un centrage du rayon 2 cm au dessus du bord supérieur de la symphyse pubienne. Par rapport au centrage sur le bord supérieur de la symphyse pubienne, le centrage du rayon incident sur l'ombilic lors de la réalisation des radiographies standards était associée à une augmentation de l'erreur de mesure (écart-type des différences entre les mesures).

D'autre part, nous n'avons pas montré de biais entre les mesures de la hauteur de l'interligne articulaire sur des radiographies réalisées avec une rotation interne des membres inférieurs du patient de 15° et celles des radiographies réalisées avec une rotation interne de 5°. Cependant, une modification de la rotation interne des membres inférieurs de 10° était associée à une augmentation de l'erreur de mesure.

Enfin, l'ensemble de cette étude nous a permis de proposer :

- 1) de réaliser les radiographies du bassin avec une rotation interne des membres inférieurs standardisée à 15° et un centrage du rayon incident à 2 cm au dessus du bord supérieur de la symphyse pubienne
- 2) un seuil de progression radiologique de la coxarthrose de 0,50 mm.

5-2-2/ Article 2: Measurement of joint space width in hip osteoarthritis: influence of joint positioning and radiographic procedure.

G-R Auleley¹, A Duche¹, J-L Drape², M Dougados¹, P Ravaud¹

1- Clinique de Rhumatologie, Hôpital Cochin, Paris, France.

2- Service de Radiologie, Hôpital Ambroise Paré, Paris, France.

Rheumatology In Press.

Joint space width measurement in hip osteoarthritis.

Measurement of joint space width in hip osteoarthritis: influence of joint positioning and radiographic procedure

Original paper

Guy-Robert AULELEY¹, MD ; Agnès DUCHE¹, MD ; Jean-Luc DRAPE², MD ; Maxime DOUGADOS¹, MD ; Philippe RAVAUD³, MD, PhD.

1- Université Paris V- René Descartes, Faculté de Médecine Cochin Port-Royal and Clinique de Rhumatologie, Hôpital Cochin, AP-HP, Paris, France.

2- Université Paris V- René Descartes, Faculté de Médecine Cochin Port-Royal and Service de Radiologie B, Hôpital Cochin, AP-HP, Paris, France.

3- Université Paris VII, Faculté de Médecine Xavier Bichat and Departement de Biostatistique et d'Epidémiologie Clinique, Hôpital Bichat-Claude Bernard, AP-HP, Paris, France.

Correspondence to: Dr Ph Ravaud, Clinique de Rhumatologie, Hôpital Cochin, 27 rue du faubourg Saint-Jacques, 75014 Paris, France.

Telephone number: 00 33 1 58 41 30 57

Fax number: 00 33 1 43 25 78 84

Summary: *Objectives:* We assessed the influence of patient positioning and radiographic procedure, and defined the smallest detectable difference (SDD) in hip OA. *Methods:* OA hip patients each had a standardized pelvic radiograph and, five minutes later, a modified pelvic radiograph with feet 5° internally rotated (part 1), X-ray beam centered on the umbilic (part 2), or another standardized pelvic radiograph (part 3). *Results:* Corresponding mean differences in joint space width (JSW) measurements (limits of agreement) (mm) between views were +0.03 (-0.53 to +0.59), -0.31 (-1.15 to +0.53), and 0.02 (-0.48 to +0.44). Both views statistically differed on mean JSW in part 2 ($p=1.6 \times 10^{-4}$), but not in part 1 ($p=0.375$) or in part 3 ($p=0.580$). SDD estimate was 0.46 mm *Conclusions:* Modifying the X-ray beam and foot rotation increases variability in JSW measurements. Use of urograms to evaluate radiological progression should be avoided. A change greater than 0.46 mm could define radiological hip OA progression.

Keywords: Hip

Osteoarthritis

Smallest detectable difference

Joint space width

Radiographic measurement

The plain radiograph is a readily available, safe, and cost-effective tool for diagnosing and assessing the progression of osteoarthritis (OA) (1,2). In both epidemiological studies and clinical trials, the best established method of assessing OA progression is by cartilage loss as measured by joint space width (JSW) at the narrowest point on plain radiographs (2,3).

Radiographic OA progression is slow (4), varying from 0.1 to 0.6 mm/year (5-7).

However, reliable assessment of OA progression requires obtaining reproducible radiographs on successive visits (8,9). Numerous sources of variability can modify JSW measurements and compromise their reliability (8-11). Such variability has been widely studied in the knee. Foot rotation has been found to be related to measurement error (10) and to significant change in JSW measurements compared to ideal joint positioning (11). On the other hand, changes induced by an X-ray beam directed below the centre of the joint space do not appear significant (10).

Such changes have been less studied in the hip although differences in patient positioning and radiographic procedure remain in the assessment of radiological hip OA. Firstly, pelvic or hip radiographs can be performed with the patient's feet internally rotated from 5° to 25° (2,12-18). Secondly, hip JSW is sometimes measured on intravenous urograms (19-23) in which the X-ray beam is centered approximately 10 cm higher than for standard anteroposterior pelvic radiographs. Although the minimum joint space in the hip may be altered by modifying foot rotation (24) or X-ray beam centering (16), the effects of these modifications on JSW measurements in hip OA are largely unknown. Thus, the aims of this study were 1) to assess the influence of internal foot rotation and X-ray beam centering on JSW measurements and 2) to determine a cutoff that allows a definition of radiological hip OA progression beyond the variability related to radiographic procedure and measurement process.

Patients and methods

Patients

Patients of both sexes with symptomatic hip OA attending the Department of Rheumatology, and willing to participate were included in this study. Hip OA was determined by the presence of pain and at least two of the following factors: radiographic osteophytes, radiographic joint space narrowing and an $ESR \leq 20$ mm/h (25). Patients were excluded if they had secondary hip OA defined by a past history of hip fracture, or the presence of calcium deposition in the hip joint, osteonecrosis, inflammatory rheumatic diseases, Paget's disease, infectious joint disease, or other hip diseases (25).

Radiography procedure

Pelvic radiographs were performed in three successive groups of twenty consecutive patients. Each patient had two pelvic radiographs taken five minutes apart. Radiographs of a given patient were taken by the same technician, in a single radiology unit, using a fast screen, cassette film and a standardized procedure. The cassette was adjusted in the holder so that the midline of the film was in alignment with the midheel line. The focus to film distance was 100 cm. Anteroposterior, standardized radiographs including both hips were taken in supine position with the X-ray beam vertical, perpendicular to the table. To obtain a constant and reproducible foot position, the inner side of each foot was aligned on a study specific V-shaped positioning frame, positioned on the table, with the back of the patient's heels lying directly on the table. For these standardized radiographs, the X-ray beam was centered 2 cm above the superior aspect of the symphysis pubis and the feet were 15° internally rotated (2,12-15) using the V-shaped positioning frame.

The first part of the study evaluated the influence of internal foot rotation (foot rotation part). In this part, patients first had a standardized radiograph with the feet 15° internally rotated;

five minutes later a modified radiograph was taken with the feet 5° internally rotated using another study specific V-shaped positioning frame. The other parameters (patient positioning and radiograph procedure) were unchanged.

The second part evaluated the influence of X-ray beam centering (X-ray beam centering part). In this part, patients had a standardized radiograph with the X-ray beam centered 2 cm above the superior aspect of the symphysis pubis, five minutes later a modified radiograph was taken With the X-ray beam centered on the umbilic. The other parameters (patient positioning and radiograph procedure) were unchanged.

The third part evaluated the reproducibility of pelvic radiographs (standardized part). In this part, pelvic radiographs were performed twice for each patient using a standardized procedure.

Radiographic assessment

Both hips of a given patient were eligible. Hips with Kellgren and Lawrence grades 2 or 3 (26) were included. In each part, the two radiographs of a single patient were placed side by side on a light box and landmarks for measurements were drawn by the reader. The landmarks consisted of two points at the narrowest point, one point on the distal margin of the condylar cortex for the femoral surface and the other on the margin of the bright radiodense band of the subchondral cortex in the floor of the articular fossa for the acetabulum. The patient's identity and the type of view were then masked with adhesive tape. The patient's identification was replaced by a random code number.

Radiographs were assessed by a single reader (GRA). The interbone distance determined by the landmarks was measured in millimetres using a 0.1 mm graduated magnifying glass laid directly over the radiograph.

Statistical analysis

In order to determine the influence of foot rotation and X-ray beam centering on JSW measurements, the magnitude of changes related to each modified view was assessed applying the graphic analysis of Bland and Altman (27-29). According to this analysis, differences between pairs of JSW measurements of the same joint of a given patient were plotted against their corresponding mean including limits of agreement defined as $d_{\text{mean}} \pm 2\text{SD}$, where d_{mean} is the mean difference and SD, the standard deviation of the differences. Differences between pairs of JSW measurements of the same joint on the standardized and the modified radiographs of a given patient were plotted against their corresponding mean. For each part, a graph was plotted including limits of agreement defined as above.

The reproducibility of standardized pelvic radiographs was also assessed using the graphic analysis of Bland and Altman (27-29). Differences between pairs of JSW measurements of the same joint of a given patient were plotted against their corresponding mean including limits of agreement. Such differences between JSW measurements on standardized radiographs repeated five minutes apart reflected the variability inherent only in the radiographic procedure and the measurement process. Thus, to be considered as statistically relevant progression, any difference between serial radiographs must be greater than this variability. The smallest detectable difference (SDD) was based on reproducibility of radiographic measurement (i.e., related to repeating the standardized radiographic procedure and measurement process) in hip OA. For a risk of $\alpha=5\%$, the SDD in radiological progression in hip OA for one individual was thus defined as 2 SD of the differences between pairs of standardized radiographs (30,31).

Joint space width measurement in hip osteoarthritis.

In addition, means of pairs of JSW measurements were compared in each part of the study using the paired *t*-test. We used a 2-tailed formulation, and *p* values <0.05 were considered statistically significant.

Results

Twenty patients were included in each part of this study. Table 1 summarizes demographic and radiological characteristics of the patients.

Thirty seven hips were available for assessment in the foot rotation part of this study. An artificial hip in one patient, and one hip graded 1 on Kellgren and Lawrence criteria in two patients were excluded. Mean (SD) of JSW was 3.10 (1.01) mm on the standardized pelvic radiograph of this part. Thirty nine hips were available for assessment in the X-ray beam centering part. One hip with JSW=0 mm was excluded. Mean (SD) of JSW was 3.17 (0.81) mm on the standardized pelvic radiograph of this part. Thirty seven hips were available in the standardized part. One artificial hip in one patient, one hip with JSW=0 mm, and one hip graded 1 on the Kellgren and Lawrence criteria were excluded. Mean (SD) of JSW was 3.39 (1.02) mm in this part.

Influence of internal foot rotation

Mean (SD) JSW was 3.07 (1.01) mm in pelvic radiographs with 5° internal foot rotation, that is, of a similar magnitude to that of pelvic radiographs with 15° internal foot rotation. Both views did not statistically differ on JSW measurements ($p=0.375$). Thus, in this study, internal foot rotation at 5° did not modify JSW on pelvic radiographs performed in supine position. The difference in JSW between 15° and 5° internal foot rotation was 0.03 mm on average (table 2) but varied from -0.70 mm to +0.60 mm. Absolute differences in JSW measurements ≥ 0.50 mm were observed in 4 of 37 cases between the two views. No specific relationship existed between the difference and the mean (figure 1A).

Influence of X-ray beam centering

On radiographs performed with the X-ray beam centered on the umbilic, mean (SD) JSW was 3.48 (0.90) mm. JSW increased on average by 10.26% (SD: 12.99%) on this view compared with the standardized pelvic radiographs. Both views statistically differed on JSW measurements ($p=1.6 \times 10^{-4}$). The difference in JSW between these views was -0.31 mm on average (table 2) but varied from -1.60 mm to +0.60 mm. All but 9 points were found to be under 0. Absolute differences in JSW measurements ≥ 0.50 mm were observed in 10 of 39 cases between the two views. No specific relationship existed between the difference and the mean (figure 1B).

Reproducibility of pelvic radiographs and smallest detectable change

JSW measurements did not differ ($p=0.580$) between the two standardized radiographs (mean \pm SD: 3.39 ± 1.02 and 3.41 ± 1.02 mm, respectively). The absolute difference in JSW between repeated radiographs was 0.02 mm on average (table 2) but varied from -0.40 mm to +0.80 mm. Absolute differences in JSW measurements ≥ 0.50 mm were observed in 1 of 37 cases between standardized repeated radiographs. No specific relationship existed between the difference and the mean (figure 1C). The estimated radiological progression in hip OA as defined by the smallest detectable difference for one individual, that is, twice the SD of the differences between standardized radiographs, was 0.46 mm.

Discussion

In this study, modifying foot rotation by 10° did not significantly change JSW on pelvic radiographs. In contrast, centering the X-ray beam on the umbilic instead of on the superior aspect of the symphysis pubis resulted in an average increase in JSW of about 10%. Radiological progression based on analysis of radiographic reproducibility was found to be 0.46 mm in this population, using a defined standardized radiographic procedure and measurement process, with this reader and measurement instrument.

Differences in JSW measurements related to modification of foot rotation were insufficient (0.03 mm) to be detectable with the measuring instrument we used. Moreover, the effect of foot rotation seems less in hip than in knee joint (10,11). It is thought that internal foot rotation might apply a load to the articular hip surfaces (14), possibly through tightening of the powerful lateral rotator muscles of the hip joint that drive the femoral head internally into the acetabulum (24). Our findings in this study, suggest that a 10° difference in internal foot rotation does not induce bias in JSW measured on pelvic radiographs of patients with hip OA, as assessed by the mean difference. However, we cannot exclude that a greater difference in foot rotation might induce a significant bias as suggested by Buckland-Wright (24). This author studied nine cadavers and reported that the joint space of the hip was narrower with radiographs obtained with the foot internally rotated than with the foot in the neutral position (24). However, our results also suggest that a 10° difference in internal foot rotation increases the variability in JSW measurement as assessed by the SD of the differences (see figures 1A and 1C, table 2).

More interesting was the effect on JSW measurements of centering the X-ray beam on the umbilic. In the knee, displacement of the X-ray beam direction led to conflicting results

(10,11). In the hip, however, the results of Conrozier et al (16) would suggest that such a displacement of the X-ray beam from the hip joint to 2 cm above the symphysis pubis can influence JSW measurement. Nevertheless, their study did not directly compare the two different patient groups, one having hip radiographs and the other pelvic radiographs. In this study, we evaluated the effect of two X-ray beam centerings, both away from the hip joint, on the same patient group. The increase in JSW on radiographs performed with the X-ray beam centered on the umbilic compared to the standardized pelvic radiograph may be due to modifications induced by the downward inclination of the X-ray exploring the joint space.

Our study assessed the reproducibility of the standardized pelvic radiograph in patients in stable condition (two radiographs five minutes apart). The cutoff observed in this study (0.46 mm) is smaller than those reported in the knee (i.e., 0.62 mm and 0.55 mm) (30,31). Possible greater variation in the radiographic process in the knee (11) may explain this difference. This cutoff is indeed close to that (0.56 mm) reported by Dougados et al (32) for hip OA despite the use of a different approach (two radiographs one year apart).

A limitation of this study is that the radiographs were performed in the supine position. Although JSW measured on standing and supine knee radiographs differ (33-36), results in the hip remain contradictory (16,37,38). However, we previously reported no significant difference in JSW measurement between the two views. In addition, technically better images may be obtained with supine than with standing pelvic radiographs, particularly in obese patients. Another limitation is that we did not compare the farthest possible degrees of rotation used to radiograph the hip, i.e., 5° and 25°. Such a range of rotation may result in significant differences in JSW measurements. However, most guidelines recommend 15° internal foot rotation to radiograph the hip joint (8,12,14,15).

In longitudinal studies, reproducible radiographs taken at different times allow reliable assessment of OA progression. For this reason, the sources of variability in JSW measurement must be minimized. First, we recommend using a positioning frame for pelvic radiographs in order to obtain a fixed degree of internal foot rotation, e.g. 15° in studies measuring JSW, instead of allowing a varying degree of rotation, e.g., $15^\circ \pm 5^\circ$. Although our study does not support a statistically significant change in mean JSW measurement due to modifications in internal foot rotation, such a change remains possible (24). Moreover, our study also suggests that the difference in internal foot rotation increases the variability of JSW measurements in the hip. Thus, using a standardized positioning frame will assure that the same rotation is used in repeating radiographs in a given patient and across studies. Second, we do not recommend evaluating radiological progression in hip OA using urograms. Pelvic radiographs should be taken with the X-ray beam precisely centered on the superior aspect of the symphysis pubis.

We have assessed the influence of some sources of variability in JSW measurement in hip OA and defined a smallest detectable change in radiological hip OA progression. These results need to be confirmed. Long-term longitudinal studies using standardized procedures as recommended in this study are also needed.

References

1. Jewell FM, Watt I, Doherty M. Plain radiographic features of osteoarthritis. In Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York/ Oxford University Press, 1998:217-37.
2. Recommendations from a task force of the Osteoarthritis Research Society : Design and conduct of clinical trials in patients with osteoarthritis. *Osteoarthritis Cartilage* 1996;4:217-43.
3. Jonsson K, Buckwalter K, Helvie M, Niklason L, Martel W. Precision of hyaline cartilage thickness measurements. *Acta Radiol* 1992;33:234-9.
4. Spector TD, Dacre JE, Harris PA, Huskisson EC. Radiological progression of osteoarthritis: an 11 year follow up study of the knee. *Ann Rheum Dis* 1992;51:1107-10.
5. Dougados M, Gueguen A, Nguyen M, *et al.* Longitudinal radiologic evaluation of osteoarthritis of the knee. *J Rheumatol* 1992;19:378-84.
6. Lequesne M. Quantitative measurements of joint space during progression of osteoarthritis: "Chondrometry". In: Kuettner KE, Goldberg V, eds. *Osteoarthritic disorders*. Illinois: American Academy of Orthopaedic Surgery, 1995:427-44.
7. Kirwan JR, Cushnaghan J, Dacre J, *et al.* Progression of joint space narrowing in knee [abstract]. *Arthritis Rheum* 1992;35 (suppl.):B103.
8. Buckland-Wright JC . Quantitative radiography of osteoarthritis. *Ann Rheum Dis* 1994;63: 268-75.
9. Ravaud Ph, Dougados M. Radiographic assessment in osteoarthritis. *J Rheumatol* 1997;24:786-91.

10. Lynch JA, Buckland-Wright C, MacFarlane DG. Precision of joint space width measurement in knee osteoarthritis from digital image analysis of high definition macroradiographs. *Osteoarthritis Cartilage* 1993;1:209-18.
11. Ravaut Ph, Auleley G-R, Chastang Cl, *et al.* Knee joint space width measurement : An experimental study of the influence of radiographic procedure and joint positioning. *Br J Rheumatol* 1996;35:761-6.
12. Buckland-Wright C, Scott Jr WW, Peterfy C. Radiographic imaging techniques. In Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York/ Oxford University Press, 1998:572-4.
13. Buckland-Wright C. Protocols for radiography. In Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York/ Oxford University Press, 1998:578-80.
14. Dieppe P, Altman RD, Buckwalter JA, *et al.* Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. In : Kuettner KE, Goldberg VM, eds. *Osteoarthritis disorders*. Rosemont, Il : American Academy of Orthopaedic Surgeons, 1994:481-96.
15. Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. *Osteoarthritis Cartilage* 1995;3:73-7.
16. Conrozier T, Lequesne MG, Tron AM, Mathieu P, Berdah L, Vignon E. The effects of position on the radiographic joint space in osteoarthritis of the hip. *Osteoarthritis Cartilage* 1997;5:17-22.
17. Hansson G, Jerrer R, Sanders SM, Wallin J. Radiographic assessment of coxarthrosis following slipped capital femoral epiphysis : A 32-year follow-up study of 151 hips. *Acta Radiol* 1993;34:117-23.
18. Resnick D, Niwayama G. *Diagnosis of bone and joint disorders*. Vol I. Philadelphia : WB Saunders, 1988/40.

19. Croft P, Cooper C, Wickham C, Coggon D. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol* 1990;132: 514-22.
20. Croft P, Cooper C, Wickham C, Coggon D. Osteoarthritis of the hip and acetabular dysplasia. *Ann Rheum Dis* 1991;50:308-10.
21. Croft P, Cooper C, Wickham C, Coggon D. Is the hip involved in generalized osteoarthritis? *Br J Rheumatol* 1992;31:325-8.
22. Lau EMC, Lin F, Lam D, Silman A, Croft P. Hip osteoarthritis and dysplasia in chinese men. *Ann Rheum Dis* 1995;54:965-9.
23. Smith RW, Egger P, Coggon D, Cawley MID, Cooper C. Osteoarthritis of the hip joint and acetabular dysplasia in women. *Ann Rheum Dis* 1995;54:179-81.
24. Buckland-Wright C. Quantitation of radiographic changes. In Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York/ Oxford University Press, 1998:459-72.
25. Altman RD. The classification of osteoarthritis. *J Rheumatol* 1995;(suppl 53)22:42-3.
26. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;16:494-502.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i:307-10.
28. Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. *Int J Epidemiol* 1995;24(suppl 1):S7-S14.
29. Bland JM, Altman DG. Measurement error. *BMJ* 1996;313:744.
30. Ravaut Ph, Giraudeau B, Auleley G-R, *et al.* Variability in knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis* 1998;57:624-9.

31. Ravaud Ph, Giraudeau B, Auleley G-R, Edouard-Noël R, Dougados M, Chastang CI. Smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. *J Clin Epidemiol* 1999;52:1225-30.
32. Dougados M, Gueguen A, Nguyen M, *et al.* Radiological progression of hip osteoarthritis : definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996;55:356-62.
33. Altman RD, Fries JF, Bloch DA, *et al.* Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987;30:1214-25.
34. Ahlback S. Osteoarthritis of the knee: a radiographic investigation. *Acta Radiol* 1968;277 (suppl):7-72.
35. Leach RE, Gregg T, Siber FJ. Weight-bearing radiograph in osteoarthritis of the knee. *Radiology* 1970;97:265-8.
36. Buckland-Wright JC, Macfarlane DG, Lynch JA, Jasani MK, Bradshaw CR. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. *Ann Rheum Dis* 1995;54:263-8.
37. Auleley G-R, Rousselin B, Ayrat X, Edouard-Noël R, Dougados M, Ravaud Ph. Osteoarthritis of the hip: agreement between joint space width measurements on standing and supine conventional radiographs. *Ann Rheum Dis* 1998;57:519-23.
38. Scott DL, Gishen P. Radiological assessment of hip osteoarthritis. *Lancet* 1999;353:87-8.

Legends of figure

Figure 1 (A) : Differences in the measurement of joint space width (in mm) between the standardized pelvic radiographs performed with 15° internal foot rotation and pelvic radiographs performed with 5° internal foot rotation (15° minus 5°) plotted against mean joint space width measurement. (B) : Differences in the measurement of joint space width (in mm) between the standardized pelvic radiographs performed with the X-ray beam centered on 2 cm above the superior aspect of the symphysis pubis and radiographs performed with the X-ray beam centered on the umbilic (pubis minus umbilic) plotted against mean joint space width measurement. (C) : Differences in the measurement of joint space width (in mm) between two standardized pelvic radiographs performed with 15° internal foot rotation and the X-ray beam centered on 2 cm above the superior aspect of the symphysis pubis plotted against mean joint space width measurement. Limits of agreement correspond to $\text{mean} \pm 2\text{SD}$.

figure 1

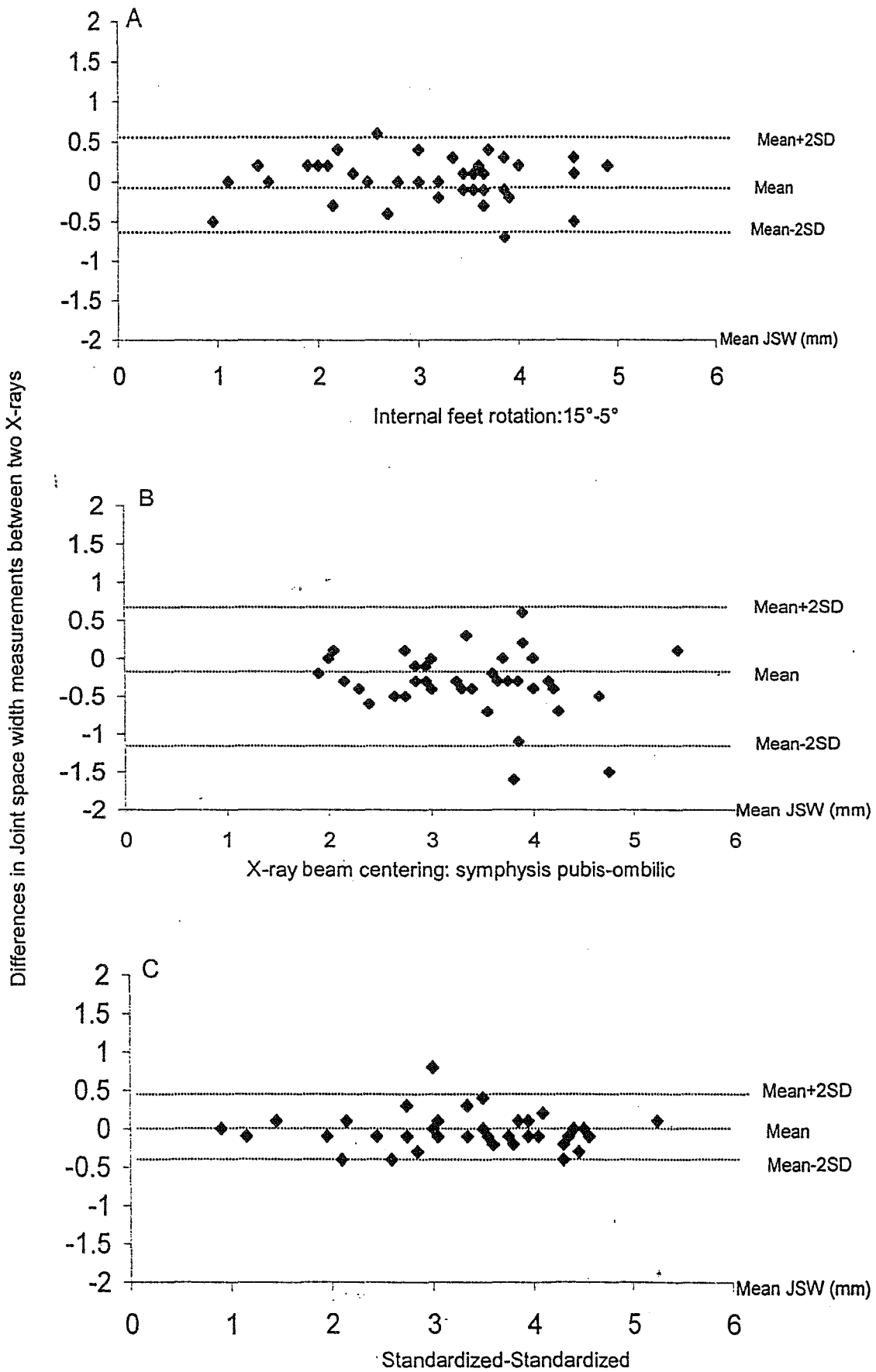


Table 1 : Demographic and radiological characteristics of patients with hip osteoarthritis assessed for the influence of X-ray beam centering and internal foot rotation on joint space width measurement, and reproducibility of standardized pelvic radiographs.

Part	Foot rotation	X-ray beam centering	Standardized
Number of patients	20	20	20
Age, mean (SD)	67.8 (10.11)	70.7 (12.59)	64.6 (10.20)
range, y	52-83	50-87	50-84
No., female	11	15	11
Height, mean (SD)	162.4 (8.70)	162.6 (9.02)	163.1 (8.13)
range, cm	147-176	143-176	146-180
Weight, mean (SD)	74.63 (16.41)	77.0 (14.45)	77.79 (11.89)
range, kg	45-107	49-101	58.5-102
Number of hips	37	39	37
Kellgren and Lawrence criteria			
2	18	17	13
3	19	22	24

Table 2 : Mean differences in joint space width measurement (JSW) (in mm), between a standardized pelvic radiograph and either pelvic radiograph with 5° internal foot rotation (foot rotation part) or X-ray beam centering on the umbilic (X-ray beam centering part), and between two standardized pelvic radiographs (standardized part). Differences are calculated on JSW measurements as values on both standardized radiographs, and as values on standardized radiographs minus modified radiographs when appropriate.

Part	Number	Mean of differences	Standard deviations	95% confidence intervals	Limits of agreement
Foot rotation	37	+0.03	0.28	-0.06; +0.12	-0.53 to +0.59
X-ray beam centering	39	-0.31	0.42	-0.44; -0.18	-1.15 to +0.44
Standardized	37	0.02	0.23	-0.07; +0.05	-0.48 to +0.44

5-3)- ETUDE DE L'INFLUENCE DE L'ORDRE DE LECTURE DES RADIOGRAPHIES ET DU MARQUAGE DU SITE DE MESURE DANS LES ETUDES LONGITUDINALES

5-3-1/ Etat des connaissances et objectifs de la recherche

Dans notre troisième travail, nous nous sommes intéressés à la technique de lecture des radiographies dans les études longitudinales.

A ce jour, il n'existe pas de consensus sur ces techniques (lecture en ordre chronologique, marquage du site de mesure). Nous avons également voulu évaluer l'influence de ces méthodes de lecture sur le calcul du nombre de sujets nécessaires à inclure dans les études longitudinales.

Certains auteurs utilisent le marquage du site de mesure sur l'interligne articulaire pour évaluer la progression de l'arthrose (*Altman 1987*) (*Ravaud 1998*), alors que d'autres auteurs ne le font pas (*Dieppe 1997*). Récemment, le marquage du site de mesure a été récemment recommandé par l'OARS (*Altman 1996*).

Par ailleurs, l'ordre dans lequel les radiographies successives d'un même doivent être lues peut varier selon les études (*Altman 1987*) (*Ledingham 1993*) (*Dieppe 1994*) (*Speake 1995*) (*Dougados 1996*) (*Ravaud 1998*). Dans l'évaluation de la progression de l'arthrose, les radiographies peuvent être groupées par patient et lues en ordre chronologique (*Spector 1992*) (*Dieppe 1993*), ou groupées par patient et lues en aveugle de la chronologie (*Altman 1987*) (*Dougados 1996a*) (*Ravaud 1996c*), ou encore isolément, c'est à dire en aveugle de l'identité du patient et de la chronologie (*Ledingham 1993*). Cependant, chacune de ses méthodes de lecture peut avoir des avantages et des inconvénients théoriques. Par exemple, la lecture des radiographies groupées par patient, chronologiquement ordonnées peut apporter plus d'information au lecteur que les deux autres procédures et ainsi aider à réduire l'erreur de mesure due à la qualité des radiographies.

Le choix d'une technique de lecture peut avoir une influence importante. Ainsi, chez les malades atteints de la polyarthrite rhumatoïde, une différence de 12 % à 38 % dans le nombre de sujets nécessaires à inclure peut être observée dans un essai thérapeutique selon que l'on évalue la progression radiologique des lésions articulaires par couples de radiographies chronologiquement ordonnées ou non (*Ferrara 1997*).

A notre connaissance, il n'existe pas d'étude ayant évalué l'influence de ces différentes techniques de lecture sur la mesure de l'interligne articulaire dans la coxarthrose.

Nous avons décidé de réaliser ce travail pour comparer ces différentes techniques de mesure et pour évaluer leur influence relative sur le nombre de sujets nécessaires à inclure dans les études longitudinales.

Nous avons donc comparé les différentes techniques de lecture suivantes :

- Radiographies groupées par patient, lues par couples (J fin d'étude-J0), dans l'ordre chronologique.
- Radiographies groupées par patient, lues par couples (J fin d'étude-J0), en aveugle de la chronologie
- Radiographies lues chacune isolément, en aveugle de l'identité et de la chronologie.

Nous avons utilisé le score de Kellgren et Lawrence et le pincement de l'interligne articulaire et la mesure de la hauteur de l'interligne articulaire pour évaluer la progression radiologique de la coxarthrose sur trois ans.

La progression de la coxarthrose a été évaluée par le score de Kellgren et Lawrence, le score de pincement articulaire et la mesure de la hauteur de l'interligne articulaire sur 3 ans.

Sur les radiographies lues isolément, la hauteur de l'interligne était mesurée après marquage du site de mesure dans un premier temps puis, sans marquage dans un deuxième temps.

Nous avons utilisé des méthodes statistiques descriptives pour comparer les différentes techniques de lecture (histogrammes de la progression, analyse en composantes principales, coefficient de corrélation intra-classe, réponse standardisée moyenne).

Dans un second temps, l'influence de ces différentes techniques de lecture sur le calcul du nombre de sujets nécessaires dans les études longitudinales a été évalué pour la mesure de la hauteur de l'interligne articulaire. Pour cela, nous avons comparé leurs réponses standardisées moyennes.

Nous avons montré que la progression des grades du score de Kellgren et Lawrence et du score de pincement de l'interligne dépendait fortement de la technique de lecture utilisée. Nous avons également montré que la lecture des radiographies groupées par patient dans l'ordre chronologique n'apportait pas plus d'information sur la mesure de la hauteur de l'interligne articulaire que les autres techniques de lecture.

Nous avons également montré que dans les études longitudinales mesurant l'interligne avec marquage du site de mesure, il faudrait inclure 10% moins de sujets si les radiographies sont lues groupées par patient, en aveugle de la chronologie plutôt qu'isolément. D'autre part,

il faudrait inclure 10% moins de sujets si les mesures de l'interligne sont obtenues avec marquage plutôt que sans marquage sur des radiographies lues isolément.

A partir de ces résultats, nous avons pu recommander d'évaluer la progression de l'interligne articulaire sur des radiographies groupées par patient et lues en aveugle de l'ordre chronologique et avec marquage du site de mesure.

5-3-2/ Article 3: Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies.

G-R Auleley¹, B Giraudeau², M Dougados¹, P Ravaud¹

1- Clinique de Rhumatologie, Hôpital Cochin, Paris, France.

2- Centre de Recherche Clinique, Faculté de Médecine, Tours, France.

Ann Rheum Dis 2000; 59: 422-27.

EXTENDED REPORTS

Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies

Guy-Robert Auleley, Bruno Giraudeau, Maxime Dougados, Philippe Ravaud

Abstract

Objective—To compare radiographic reading procedures and evaluate their impact on sample size in hip osteoarthritis (OA) longitudinal studies.

Methods—Pelvic radiographs performed twice, three years apart, in 104 patients with hip OA were read by a single reader using the Kellgren and Lawrence system, joint space narrowing scale, and joint space width (JSW). Reading procedures were (a) films read as single radiographs, (b) films grouped by patient but read in random order, (c) films grouped by patient and chronologically ordered, all with landmarks for JSW measurements, (d) films read as single radiographs, without landmarks for JSW measurements. JSW was measured at the narrowest point with a 0.1 mm graduated magnifying glass.

Results—More Kellgren and Lawrence or joint space narrowing grades were modified respectively with the single (42% and 37%) than with the paired (32% and 23%) or chronologically ordered (34% and 29%) reading procedures. Variability of JSW progression was principally related to mean progression (88.3%) and landmarks (almost 10%). Standardised response means were -0.71 with the paired reading procedure with landmarks, -0.68 with the single reading procedure with landmarks, -0.65 with the single reading procedure without landmarks. With landmarks, 10% more patients would be needed using single than paired reading. Using single reading, 10% more patients would be needed without landmarks than with landmarks.

Conclusion—Kellgren and Lawrence grading seems to be influenced by the reading procedure, as is joint space narrowing grading, for assessing hip OA. Paired reading procedure with landmarks for JSW should be recommended in longitudinal studies.

(Ann Rheum Dis 2000;59:422-427)

osteoarthritis (OA).¹⁻³ Depending on the joint studied, several indices are currently used for assessing radiological progression of OA, including individual radiographic features, composite indices, and quantitative measures.⁴ The Kellgren and Lawrence grading system is often used despite its limitations.⁵⁻⁷ Joint space narrowing, recorded both as measurements of interbone distance or by visual grading, is presently one of the most common variables recommended.^{1,2} However, methodological problems remain when using this index. Firstly, when measuring joint space width (JSW) for assessing OA radiological progression, landmarks for measurements may be drawn on radiographs^{7,8} or not.^{9,10} The implications of drawing landmarks have not yet been studied. Secondly, it has been recommended that the date and identification of the patient of radiographs should not be known when assessing the progression of OA,¹¹ and the type of blinding differs in studies of OA progression.^{7-10,12,13} Thus radiographic OA progression might be assessed with the reader aware of neither the patient's identity nor the chronological order of the radiographs (single reading procedure), aware of the patient's identity but unaware of the chronological order of the radiographs (paired reading procedure), or aware of both the patient's identity and the chronological order of the radiographs (chronologically ordered reading procedure). Yet these radiographic reading procedures seem to have different implications on the power of studies.^{14,15} In clinical trials on rheumatoid arthritis, 11.8% and 38% more patients would be needed to detect the same progression difference of joint damage and joint erosion scores, respectively, when reading paired films chronologically ordered rather than reading paired films in random order.¹⁴

These different reading procedures have not been compared in the assessment of OA progression. Additionally, the impact on the design of OA studies of blindness or drawing landmarks for measuring radiological progression has not been evaluated. The aims of this study were (a) to compare different reading procedures and (b) to evaluate the impact of these reading procedures and using landmarks to measure JSW on sample size, for longitudinal evaluation of hip OA.

Faculté de Médecine
Cochin Port-Royal and
Clinique de
Rhumatologie, Hôpital
Cochin, AP-HP, Paris,
France
G-R Auleley
M Dougados
Ph Ravaud

Centre de Recherche
Clinique, Faculté de
Médecine, Tours,
France
B Giraudeau

Correspondence to:
Dr Ph Ravaud, Clinique de
Rhumatologie, Hôpital
Cochin, 27 rue du faubourg
Saint Jacques, 75014 Paris,
France
Email: philippe.Ravaud@
cch.ap-hop-paris-fr

Accepted for publication
25 February 2000

Structural morphological changes on radiographs are considered the primary outcome variables for assessing the progression of

Patients and methods

PATIENTS

A sample of 104 patients fulfilling the American College of Rheumatology clinical and radiographic criteria for the diagnosis of hip OA including hip pain with at least two of the following: joint space narrowing, osteophytes, and erythrocyte sedimentation rate <20 mm/1st h,¹⁶ were selected from a three year randomised trial, controlled trial. In this randomised trial, other inclusion criteria were: age between 50 and 75 years, presence of daily pain for at least one month in the past three months, absence of secondary hip OA (presence or past history of hip fracture, inflammatory rheumatic disease, osteonecrosis, Paget's disease, etc),¹⁷ a JSW larger than 1 mm at the narrowest point, absence of medial or axial femoral head migration, or both, on radiographs, obtaining written informed consent. To enter our study, patients were selected using the following criteria: available radiographs for three years and absence of hip prosthesis on the target joint.

RADIOGRAPHS

Each patient had plain pelvic radiograph at entry into the study and at three years. Anteroposterior radiograph was performed with the patient standing on both legs. The patient's feet were $15^\circ \pm 5^\circ$ internally rotated. The x ray beam was horizontal, perpendicular to the table. The source to film distance was 100 cm. Thus 208 radiographs—that is, two radiographs for each patient were obtained.

RADIOGRAPHIC ASSESSMENT

Radiographs were assessed by one reader (GRA). For quantitative measurement, the interbone distance at the narrowest point was measured in millimetres using a 0.1 mm graduated magnifying glass laid directly over the radiograph. In addition, joint space narrowing was graded 0–3 using a radiographic atlas.¹⁸ Overall severity of OA was graded using the Kellgren and Lawrence grading scale¹⁹ defined as follows: 0 = normal; 1 = doubtful narrowing of joint space and possible osteophytic lipping; 2 = definite osteophytes and possible narrowing of joint space; 3 = moderate multiple osteophytes, definite narrowing of joint space, some sclerosis, and possible deformity of bone contour; 4 = large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone contour.

READING PROCEDURES

Four reading sessions were performed one week apart. At each session the reader used a different reading procedure and was unaware

of the results of the previous sessions. Radiographs were read as single, paired, or chronologically ordered depending on whether the patient's identity or time sequence, or both, were known or unknown (table 1).

For the first three reading procedures, hip OA progression was assessed using the Kellgren and Lawrence grading scale, joint space narrowing grading scale, and JSW measurement. JSW was measured with landmarks. Landmarks were drawn before the first reading session. To draw landmarks, the two radiographs of each patient were placed side by side on a light box and landmarks were immediately drawn. The landmarks consisted of two points, one on the distal margin of the condylar cortex for the femoral surface and the other on the margin of the bright radiodense band of the subchondral cortex in the floor of the articular fossa for the acetabulum.

For the single reading procedure all 208 radiographs were read as single. Patient's identity and date of radiographs were masked with adhesive tape using a randomisation list.

For the paired reading procedure, radiographs performed at entry and at three years were grouped by patient. Only the date of radiographs was masked with adhesive tape using another randomisation list. Pairs of radiographs of the same patient (n=104) were read side by side in random order, the chronological order being unknown to the reader.

For the chronologically ordered reading procedure, radiographs were grouped by patient and read by pairs chronologically ordered, the chronological order being known to the reader.

After these first three reading sessions, landmarks were erased.

For the single reading procedure without landmarking the 208 radiographs were read blindly as single and JSW measurements were made without previous landmarks.

STATISTICAL ANALYSIS

To compare the reading procedures when assessing hip OA progression by Kellgren and Lawrence and joint space narrowing grading scales, we constructed histograms of changes in grades (values at three years minus values at entry).

The effects of reading procedures when assessing hip OA progression by JSW measurements were assessed using the same methods. Comparisons of paired, chronologically ordered, single reading procedures with landmarks and single reading procedure without landmarks were performed using descriptive statistics. These included means and standard deviations of the differences between the values of JSW measured on the last and the first radiographs of the same patient (values at three years minus values at entry) and histograms showing the progression of JSW.

A principal component analysis was also performed. The object of this analysis was to take the four values of change in JSW measurements (that is, one value for each reading procedure) for each patient and to find combinations of these to produce mutually uncorrelated indices (named principal compo-

Table 1 Sessions and procedures of reading

Reading sessions	Landmarks	Patient identity	Time sequence	Reading procedures
First	Present	Blinded	Blinded	With landmarks
Second	Present	Known	Blinded	Single
Third	Present	Known	Known	Paired
				Chronologically ordered
Fourth	Absent	Blinded	Blinded	Without landmarks
				Single

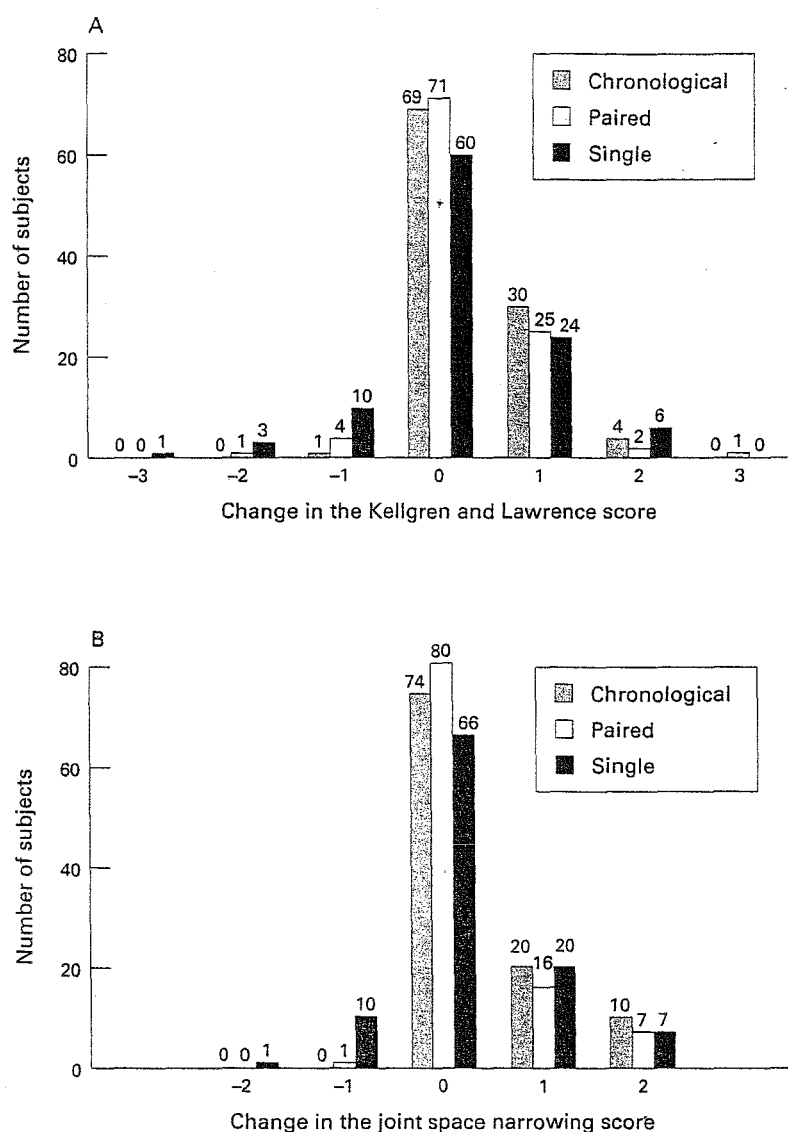


Figure 1 Distribution of changes in Kellgren and Lawrence score and joint space narrowing score using three reading procedures in 104 patients with hip osteoarthritis.

nents). This lack of correlation between principal components then allowed better understanding of the differences emerging from the four reading procedures. Moreover, principal components were ordered so that the first component explained the largest amount of total variability, the second component explained the second largest amount of total variability, and so on. Estimations of intraclass correlation coefficients (ICCs) of pairs of reading procedures were derived in the framework of a two way random effect model. Approximate 95% confidence intervals were estimated by the Fleiss and Shrout result.²⁰

Table 2 Progression of joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis followed up for three years

	Reading procedure			
	Chronologically ordered with landmarks	Paired with landmarks	Single with landmarks	Single without landmarks
Mean (mm)	-0.60	-0.59	-0.58	-0.47
Standard deviation (mm)	0.85	0.83	0.85	0.73
Standardised response mean*	-0.70	-0.71	-0.68	-0.65

*Standardised response mean is the ratio of mean and standard deviation of change.

Standardised response mean (SRM)—that is, a responsiveness statistic indicating the magnitude of change (mean difference) in comparison with the standard deviation of change,²¹ was estimated for chronologically ordered, paired, and single reading procedures with landmarks or single reading procedure without landmarks. As it has been shown that high responsive instruments lead to lower sample size requirements,²² the effects of these reading procedures on sample size requirements when using a paired Student's *t* test were evaluated by comparing the corresponding SRM.

Results

At baseline, patients had a mean (standard deviation) age of 62.3 (7.2) years, weight 69.9 (11.8) kg, and height 165.5 (9.1) cm.

When assessed with the single reading procedure, the Kellgren and Lawrence grade changed in 44 patients (42%) compared with change in 35 (34%) and 33 (32%) patients when assessed with the chronologically ordered or paired reading procedures, respectively (fig 1). Of note, on the Kellgren and Lawrence scale 14 patients had an improvement with the single reading procedure against only one with the chronologically ordered reading procedure. When assessed with the single reading procedure, the joint space narrowing grade changed in 38 patients (37%) compared with change in 30 (29%) and 24 (23%) patients when assessed with the chronologically ordered or paired reading procedures, respectively. Although an improvement in Kellgren and Lawrence or joint space narrowing grading scales by more than one grade (change towards a lesser grade) was rarely seen (<10%), the frequency of progression differed according to the radiographic feature and reading procedure. Improvement on the Kellgren and Lawrence and joint space grading scale was more often seen with the single reading procedure than with the other reading procedures. Except for two patients, those in whom improvement was seen on the Kellgren and Lawrence grading scale differed from those in whom improvement was seen on the joint space narrowing grading scale. For these two patients, improvement was seen only with the single reading procedure.

Table 2 and fig 2 show JSW progression according to the reading procedure. The progression was less with the single reading procedure without landmarks (-0.47 mm) than with the other reading procedures (at least -0.58 mm for the single with landmarks). Progression standard deviation was also lower with the single reading procedure without landmarks (0.73 mm) than with the other reading procedures (0.85 mm for the single with landmarks).

Figure 3 shows the results of principal component analysis. The two first principal components represented almost 98% of the total variability of JSW progression. The first principal component was responsible for 88.3% of the total variability and corresponded to the progression of JSW. The second principal

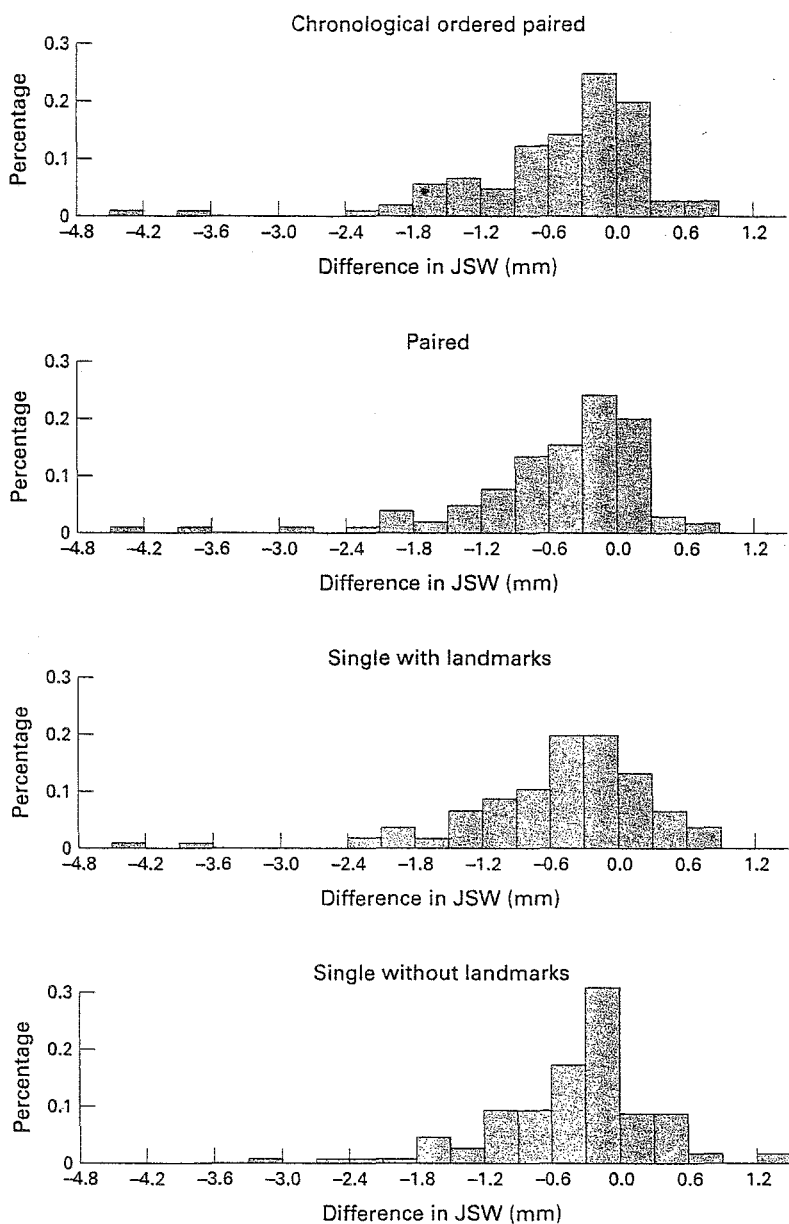


Figure 2 Distribution of the progression in joint space width (JSW) using four reading procedures in 104 patients with hip osteoarthritis.

component contrasted the reading procedures using landmarks with the procedure without landmarks and was responsible for about 10% of the total variability.

Table 3 gives details of the ICCs and their corresponding 95% confidence intervals. There was a high agreement between reading procedures performed with landmarks as ICCs reached 0.96. The agreement between each of the three reading procedures with landmarks and the one without landmarks was less, though the corresponding ICCs remained at about 0.70.

The paired reading procedure of JSW measurements with landmarks had the highest SRM (-0.71), whereas the reading procedure without landmarks had an SRM of -0.65 (table 2). The ratio between estimated sample sizes when comparing two reading procedures actually amounts to the square of the ratio of the corresponding SRM. This ratio equalled 1.10 or 1.06 when comparing the paired reading procedure with landmarks or the chronologi-

cally ordered reading procedure with landmarks with the single reading procedure with landmarks, respectively. Therefore, 10% or 6% more patients would be needed when using the single reading procedure with landmarks than when using the paired reading procedure with landmarks or the chronologically ordered reading procedure with landmarks, respectively. When comparing the paired reading procedure with landmarks with the chronologically ordered reading procedure with landmarks, this ratio equalled 1.03. That is, when using the chronologically ordered reading procedure with landmarks, 3% more patients would be needed.

The ratio between estimated sample sizes when using the single reading procedure with and the single reading procedure without landmarks equalled 1.10. That is, 10% more patients would be needed than when using the single reading procedure without landmarks in longitudinal studies.

Discussion

In this study we assessed the radiographic progression of hip OA on pelvic plain radiographs taken three years apart in 104 patients with hip OA. Single, paired, and chronologically ordered reading procedures resulted in a difference in the number of patients changing grades on Kellgren and Lawrence or joint space narrowing grading scales. The single reading procedure more often showed change on the Kellgren and Lawrence scale (42% of patients) and on the joint space narrowing scale (37% of patients) than each of the other reading procedures. Measurement of JSW progression on single radiographs without landmarks would require 10% more patients than on single radiographs with landmarks in longitudinal studies.

The data reported in this study emphasise the consequences of using insensitive methods for measuring radiological OA progression in terms of sample size and, therefore, of the cost of studies. For example, in longitudinal studies, measurement of JSW progression on single radiographs with landmarks would necessitate 10% more patients than on paired radiographs with landmarks.

Although several recommendations for assessing OA progression exist,¹⁻³ the need for standardised methodology remains a challenge.²³ The Kellgren and Lawrence grading scale has been shown to be poorly responsive in knee OA.²⁴ Clearly, the results of this study suggest that this index is also dependent on the reading procedure in assessing hip OA progression. The change in the Kellgren and Lawrence scale was seen more often with the single reading procedure than with the paired reading procedures. Such findings have been reported elsewhere for other progression scores when reading paired radiographs was compared with reading single radiographs.²⁵

Individual features, such as the joint space narrowing scale, have been presumed to be more informative in prospective studies.²⁶ Altman *et al* rated joint space narrowing on pairs of radiographs as most important in

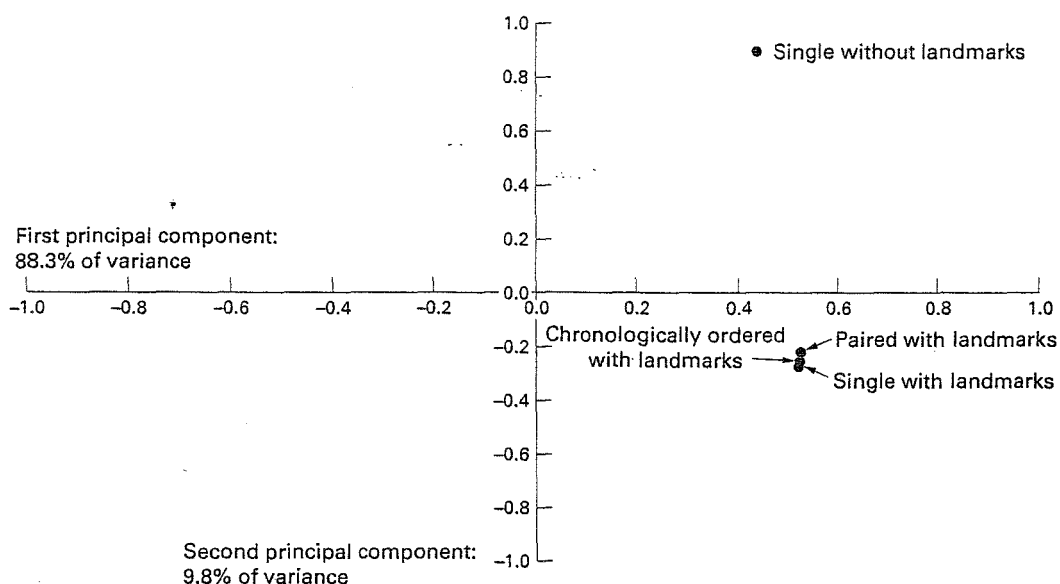


Figure 3 Principal component analysis on change in joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis. The first principal component accounts for 88.3% of the total variability. All the four coefficients are positive and nearly equal, thus allowing the assertion that this component is a measure of global time variation value. The second principal component opposes the single reading procedure without landmarks to the three other reading procedures. This component, which accounts for 9.8% of total variability, then reflects the advantage of using landmarks.

determining progression of hip OA by identifying the correct time sequence of radiographs.⁷ We did not find that the results with the joint space narrowing scale were less influenced by reading procedures than results with the Kellgren and Lawrence scale for assessing hip OA progression. Our results actually show that the change in the joint space narrowing scale is also seen more often with the single reading procedure. Differences in methodologies used in these studies may explain these contradictory conclusions. While Altman *et al* used the average score of joint space narrowing rated as a percentage of narrowing by three different readers,⁷ we, in contrast, used a single reader who graded joint space narrowing using an atlas.¹⁸

For the reader there was no advantage in having information on patient identity or even on time sequence for measuring JSW in this study. Perhaps, such information would have been important if there had been a dramatic progression in JSW. However, measuring JSW on paired films would reduce the sample size required compared with the other reading procedures with landmarks. Although the purpose of blindness in clinical studies is, as far as possible, to reduce bias introduced by the investigator's knowledge of a previous measurement for a given subject,^{27, 28} the blindness effect seems smaller with a more objective outcome measure, such as JSW measurements, than with outcome measures requiring observers'

judgment, such as Kellgren and Lawrence or joint space narrowing scales.

More interesting is the influence of drawing landmarks when assessing JSW progression. Landmarks have been recommended for measuring JSW in OA clinical trials.³ According to descriptive statistics, reading procedures with landmarks clearly seemed similar in our study. The presence of landmarks accounted for almost 10% of the variability in JSW progression, and the reading procedures with landmarks clearly contrasted with the reading procedure without landmarks. We considered SRM for comparison of the respective effect on sample size requirements because not only the SD of JSW progression but also the mean of JSW progression differed according to the reading procedure. JSW progression was more homogeneous and at the same time the progression was less with the reading procedure without landmarks than with reading procedures with landmarks. It may be that JSW measurements are not performed in the same sites with these respective reading procedures. Therefore, 10% more patients would be needed in longitudinal studies with a single reading procedure without landmarks than with a single reading procedure with landmarks.

A limitation of this study is that we did not evaluate paired and chronologically ordered reading procedures without landmarks for JSW measurements. We considered that evaluating

Table 3 Intraclass correlation coefficients for change in joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis

Reading procedure	Reading procedure		
	Paired with landmarks	Single with landmarks	Single without landmarks
Chronologically ordered with landmarks	0.963 (0.947 to 0.975)	0.958 (0.938 to 0.971)	0.698 (0.583 to 0.785)
Paired with landmarks		0.963 (0.946 to 0.975)	0.715 (0.606 to 0.798)
Single with landmarks			0.695 (0.580 to 0.782)

these reading procedures without landmarks would have a limited interest and would give results close to those with landmarks. Nor were differences between repeated readings of each procedure by the same reader assessed. For reading procedures using landmarks, these differences may be expected to be almost the same if these reading procedures are compared by pairs.

Although other studies are needed to confirm these findings, the paired reading procedure with landmarks seems to be the most appropriate procedure and should be recommended for measuring JSW progression in hip OA clinical studies.

- Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. *Osteoarthritis Cartilage* 1995;3:73-7.
- Group for the respect of ethics and excellence in science for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis* 1996;55:552-7.
- Recommendations from a task force of the Osteoarthritis Research Society: design and conduct of clinical trials in patients with osteoarthritis. *Osteoarthritis Cartilage* 1996;4:217-43.
- Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. *J Rheumatol* 1997;24:786-91.
- Spector TD, Hochberg MC. Methodological problems in the epidemiological study of osteoarthritis. *Ann Rheum Dis* 1994;53:143-6.
- Murphy Jr WA, Altman RD. Updated osteoarthritis reference standard. *J Rheumatol* 1995;22(suppl 43):56-9.
- Altman RD, Fries JF, Bloch DA, Cartens J, Cooke TD, Genant H, et al. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987;30:1214-25.
- Ravaud P, Giraudeau B, Auleley G-R, Drape J-L, Rousselin B, Paolozzi L, et al. Variability of knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis* 1998;57:624-9.
- Dieppe PA, Cushnaghan J, Shepstone L. The Bristol "OA500" study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. *Osteoarthritis Cartilage* 1997;5:87-97.
- Speake SC, Szebenyi B, Dieppe P. Quantitative serial assessment of patellofemoral joint osteoarthritis from lateral knee radiographs. *Osteoarthritis Cartilage* 1995;3(suppl A):89-96.
- Dieppe P, Altman RD, Buckwalter JA, Felson DT, Hascall V, Lohmander LS, et al. Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. In: Kuettner KE, Goldberg VM, eds. *Osteoarthritis disorders*. Rosemont, IL: American Academy of Orthopedic Surgeons, 1994:481-96.
- Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, et al. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996;55:356-62.
- Ledingham J, Dawson S, Preston B, Milligan G, Doherty M. Radiographic progression of hospital referred osteoarthritis of the hip. *Ann Rheum Dis* 1993;52:263-7.
- Ferrara R, Priolo F, Cammisa M, Bacarini L, Cesare A, Pasero G, et al. Clinical trials in rheumatoid arthritis: methodological suggestions for assessing radiographs arising from the GRISAR study. *Ann Rheum Dis* 1997;56:608-12.
- van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology* 1999;38:1213-20.
- Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505-14.
- Altman RD. The classification of osteoarthritis. *J Rheumatol* 1995;22(suppl 43):42-3.
- Altman RD, Hochberg M, Murphy Jr WA, Wolpe F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995;3(suppl A):3-70.
- Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16:494-501.
- Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* 1978;43:259-62.
- Fortin PR, Stucki G, Katz JN. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum* 1995;38:1027-30.
- Tuley MR, Mulrow CD, McMahan CA. Estimating and testing an index of responsiveness and the relationship of the index to power. *J Clin Epidemiol* 1991;44:417-21.
- Dieppe P, Brandt KD, Lohmander S, Felson DT. Detecting and measuring disease modification in osteoarthritis. The need for standardized methodology. *J Rheumatol* 1995;22:201-3.
- Ravaud P, Giraudeau B, Auleley G-R, Chastang C, Poirau-deau S, Ayrat X, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. *J Rheumatol* 1996;23:1756-64.
- Fries JF, Bloch DA, Sharp JT, McShane DJ, Splitz P, Bluhm GB, et al. Assessment of radiological progression in rheumatoid arthritis. A randomized, controlled trial. *Arthritis Rheum* 1986;29:1-9.
- Hart DJ, Spector TD. Radiographic criteria for epidemiologic studies of osteoarthritis. *J Rheumatol* 1995;22(suppl 43):46-8.
- Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1994.
- Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford: Blackwell Scientific, 1994.

5-4)- ESSAIS THERAPEUTIQUES ET COXARTHROSE: ETUDE DE L'INFLUENCE DU CRITERE DE JUGEMENT RETENU ET DE LA DUREE D'ETUDE SUR L'EFFECTIF À INCLURE

5-4-1/ Etat des connaissances et objectifs de la recherche

Dans un quatrième travail, nous nous sommes intéressés au calcul du nombre de sujets nécessaires et à la durée des essais thérapeutiques de traitements anti-arthrosiques à visée structurale, en fonction de différents critères de jugement.

De nombreuses molécules sont actuellement testées pour leurs propriétés anti-arthrosiques (*Altman 1998*). Bien qu'aucun traitement anti-arthrosique à visée structurale n'ait été formellement montré efficace pour modifier les dégradations cartilagineuses chez l'homme (*Pavelká 2000*), certaines molécules paraissent prometteuses (*Reginster 1999*) (*Dougados 2000*).

De nombreux auteurs recommandent la hanche comme l'une des articulations principales pour tester l'efficacité des traitements anti-arthrosiques à visée structurale (*Lequesne 1994*) (*Altman 1996*) (*Bellamy 1997*) (*Hochberg 1997*).

Dans de ces essais thérapeutiques, le critère de jugement principal doit être une variable morphologique (*Altman 1996*) (*GREES 1996*) (*Bellamy 1997*). Le nombre de sujets nécessaires à inclure doit être calculé pendant la phase de planification de l'essai en utilisant le critère de jugement principal (*Altman 1996*). Cependant, plusieurs problèmes restent posés concernant le calcul du nombre de sujets nécessaires à inclure dans l'évaluation de ces traitements à la hanche :

- Le choix du critère de jugement à utiliser pour calculer le nombre de sujets nécessaires,
- L'utilisation du critère de jugement comme variable quantitative, qualitative, ou de survie,
- L'influence des données manquantes dans le calcul du nombre de sujets nécessaires.

Premièrement, différents critères de jugement ont été proposés pour évaluer ces traitements. La mesure de l'interligne articulaire est actuellement le critère de jugement principal recommandé dans ces essais thérapeutiques (*Altman 1996*). Ce critère est habituellement utilisé comme variable quantitative (*Lequesne 1994a*).

Par ailleurs, l'intérêt récent accordé à la reproductibilité des critères de jugement a conduit certains auteurs à proposer un seuil de progression de l'arthrose basée sur l'erreur de mesure de l'interligne articulaire. Ce seuil permet ainsi de calculer la taille des essais thérapeutiques en utilisant la mesure de l'interligne articulaire comme variable qualitative (Ravaud 1998).

Récemment, la mise en place d'une prothèse totale de hanche a été également proposée comme critère de jugement valide pour évaluer l'efficacité des traitements anti-arthrosiques à visée structurale (Dougados 1999). Ce critère de jugement peut être utilisé comme variable qualitative ou variable censurée.

Cependant, le choix du critère de jugement et son utilisation peuvent avoir une influence sur le calcul du nombre de sujets à inclure.

Quelque soit le critère de jugement choisi, le calcul du nombre de sujets nécessaires doit être basé sur des estimations réalistes de la progression de l'arthrose, à partir des données d'études cliniques ou épidémiologiques. Cependant, ces essais pouvant durer plusieurs années, l'existence de données manquantes (perdus de vue, radiographies non faites, prothèses totales de hanche pour la mesure de l'interligne) peut concerner un nombre important de sujets. Leur traitement par des méthodes telles que le « last observation carried forward » (LOCF), consistant à remplacer la donnée manquante chez un patient donné par sa dernière observation disponible, n'apparaît pas appropriée. En effet, cette méthode peut par exemple conduire à considérer que les patients sortis prématurément d'essai pour inefficacité après 6 mois de traitement, ont un interligne inchangé.

A ce jour, l'influence du critère de jugement, de son utilisation et des données manquantes sur le calcul du nombre de sujets à inclure dans les essais thérapeutiques de traitements anti-arthrosiques à visée structurale n'est pas connue.

Dans ce travail, nous avons voulu estimer le nombre de sujets nécessaires à inclure dans de tels essais en fonction de différents critères et différentes durées d'étude.

Nous avons utilisé les données de progression de l'interligne et de la mise en place de prothèse totale de hanche de sujets recueillies dans le groupe placebo d'une étude multicentrique, prospective pendant trois ans.

Nous avons utilisé plusieurs critères prenant en compte les différents problèmes posés par le calcul du nombre de sujets nécessaires dans les essais thérapeutiques de la coxarthrose.

Les critères utilisés étaient les suivants:

- 1) Mesure de la hauteur de l'interligne articulaire au point le plus pincé. Pour ce critère, nous avons considéré qu'en cas de mise en place d'une prothèse totale de hanche, la mesure de l'interligne articulaire était une valeur manquante.

- 2) Mesure de la hauteur de l'interligne articulaire au point le plus pincé. Pour ce critère, nous avons considéré qu'en cas de mise en place d'une prothèse totale de hanche, la hauteur de l'interligne était égale à 0 mm.
- 3) Nombre de patients ayant une progression de la coxarthrose définie par un pincement de l'interligne supérieur à 0,5 mm ou la mise en place d'une prothèse totale de hanche.
- 4) Nombre de patients ayant une progression de la coxarthrose définie par un pincement de l'interligne supérieur à 0,5 mm. En cas de mise en place d'une prothèse totale de hanche, la mesure de l'interligne articulaire était considérée comme valeur manquante.
- 5) Nombre de patients nécessitant la mise en place d'une prothèse totale de hanche.

Les deux premiers critères ont été traités comme variables quantitatives, le troisième et le quatrième comme variables qualitatives, et le dernier comme variable de survie.

Nous avons pris l'hypothèse d'un essai thérapeutique avec deux bras parallèles contrôlé, contre placebo, testé des effets traitement de 50 % et de 30 % sur la progression entre les deux bras et des durées d'étude de un, de deux et de trois ans.

Selon le critère utilisé, le nombre de sujets nécessaires à inclure dans des tels essais durant 1 à 3 ans, variait de 41 à 559, pour un effet traitement de 50%.

Nous avons également montré qu'avec 4 de ces critères, le nombre de personnes-années de suivi dans les essais durant 2 ans était supérieur de 8% à 29% à celui des essais durant 1 an. Le nombre de personnes-années dans les essais durant 3 ans était supérieur d'environ 12% à 33% à celui des essais durant 1 an. Par contre, les différences en nombre de personnes-années entre les études durant 2 et 3 étaient inférieures à 10%. Enfin, en utilisant la mise en place d'une prothèse totale de hanche à 1, 2 ou 3 ans comme critère, les différences en nombre de personnes-années entre les différentes durées d'essai étaient inférieures à 10% dans tous les cas.

Ces résultats nous ont permis de proposer de prendre en compte le nombre de personnes-années de suivi et la pertinence clinique du critère retenu dans le calcul du nombre de sujets nécessaires à inclure dans les essais thérapeutiques de traitements anti-arthrosiques à visée structurale.

5-4-2/ Article 4: Sample size calculation according to study duration for different outcome measures in hip osteoarthritis clinical trials.

G-R Auleley^{1,2}, B Giraudeau^{2,3}, J-F Maillefert³, M Lequesne⁴, M Dougados¹, P Ravaud^{2,5}

1- Clinique de Rhumatologie, Hôpital Cochin, Paris, France.

2- INSERM U444, Epidémiologie et Sciences de l'information, Paris, France.

2- Centre de Recherche Clinique, faculté de Médecine, Tours, France.

3- Hôpital Général, Centre Hospitalier Universitaire, Dijon, France.

4- Collège de Médecine, Paris, France.

5- Département de Biostatistiques et d'Epidémiologie Clinique, Université Paris VII, Faculté de Médecine Xavier Bichat, Hôpital Bichat-Claude Bernard, Paris, France.

Soumis à publication.

Sample size calculation according to study duration for different outcome measures in hip osteoarthritis clinical trials.

Auleley Guy-Robert^{1,2}, Giraudeau Bruno^{2,3}, Maillefert Jean-Francis^{1,4}, Lequesne Michel⁵,
Dougados Maxime¹, Ravaud Philippe^{2,6}.

- 1- Université Paris V- René Descartes, Faculté de Médecine Cochin Port-Royal and Clinique de Rhumatologie, Hôpital Cochin, AP-HP, Paris, France.
- 2- INSERM U444, Epidémiologie et Sciences de l'Information, Paris, France
- 3- Centre de Recherche Clinique, Faculté de Médecine, Tours, France.
- 4- Hôpital Général, Centre Hospitalier Universitaire de Dijon, Dijon, France.
- 5- Collège de Médecine, Paris, France.
- 6- Université Paris VII, Faculté de Médecine Xavier Bichat and Département de Biostatistique et d'Epidémiologie Clinique, Hôpital Bichat-Claude Bernard, AP-HP, Paris, France.

Correspondence to: Pr Ph Ravaud, Département de Biostatistique et d'Epidémiologie Clinique, Hôpital Bichat-Claude Bernard, 46 rue Henri-Huchard, 75018 Paris, France.

Fax number: 00 33 1 43 25 78 84

e-mail address: philippe.ravaud@cch.ap-hop-paris.fr

Telephone number: 00 33 1 58 41 30 57

Abstract

In clinical trials of disease modifying osteoarthritis drugs, sample sizes should be calculated taking into account missing values. Supposing a 2-parallel group design in hip osteoarthritis clinical trials, we calculated sample sizes for 1, 2, and 3 year durations, using five parameters: 1) joint space width (JSW) measurements considering total hip arthroplasty (THA) as missing value for JSW (JSW1), 2) JSW measurements considering THA as JSW=0 mm (JSW2), 3) number of patients with joint space narrowing >0.5 mm or requiring THA (JSN+THA), 4) number of patients with joint space narrowing >0.5 mm considering THA as missing value for JSW (JSN), 5) time to THA (THA). With a 30% tested treatment effect, between 125 and 1851 patients would be required according to study duration and parameter. The corresponding person-years follow-up was directly related to the means of accounting for missing values. These results can help investigators planning clinical trials.

Keywords: Hip osteoarthritis, Disease modifying osteoarthritis drugs, Structural outcome measures, Sample size, Person-years, Study duration.

Running title: Sample size calculation with different outcome measures in clinical trials of hip osteoarthritis.

Introduction

Osteoarthritis (OA) is the most common form of arthritis (1). To date, no drug has been demonstrated in a published clinical controlled trial to modify the pathological structural changes of OA in humans (2). However, several agents are currently being evaluated for disease modifying OA (DMOAD) properties (3). The hip has been proposed as one of the main joints to test DMOAD (4-7).

Although both clinical and morphologic variables are currently recommended for DMOAD clinical trials, the primary outcome variables should be joint morphology measures (4-6). Joint space width (JSW) measured on radiographs seems the most sensitive outcome for assessing structural progression of OA, and is currently recommended as the primary outcome measure in such clinical trials (4-7). JSW measurements are usually reported as a continuous variable (7). However, the smallest detectable difference defining radiological progression (0.5 mm in hip OA) (8) permits its expression as a dichotomous variable (number of patients who progress) (9). On the other hand, the requirement for total hip arthroplasty (THA) may be a valid outcome measure in hip OA (4,7,10).

One fundamental question facing the investigators of a trial at the planning stage is to determine the sample size and the study duration. For DMOAD clinical trials aiming at stabilizing or repairing established OA, given that the rate of articular cartilage loss is usually considered to be slow, long study durations are usually recommended (7). The aim of this study was thus to estimate sample sizes required according to study durations of DMOAD clinical trials in hip OA for different outcome measures.

Patients and methods

Patients

The study population included the 252 outpatients of the placebo group of a multicenter, prospective, longitudinal, 3 year study, approved by the Ethics Review Board of Cochin Hospital, Paris (8). Patients fulfilled the American College of Rheumatology criteria for hip OA (11). They were 50 to 70 years old and had daily hip pain for at least one month in the past three, JSW > 1 mm at the narrowest point and gave written informed consent. Patients were not included if they had any of the following: secondary hip OA (presence or history of hip fracture, inflammatory rheumatic disease, osteonecrosis, Paget's disease, etc.) (12), medial or axial femoral head migration, or, both, on radiographs. Baseline characteristics of these patients, including Lequesne index assessing functional disability on daily activities during the previous seven days (score 0 to 24) (13) and Kellgren and Lawrence grading system for radiological OA severity (14), are shown in table 1.

Radiographs

Patients had pelvic radiographs at entry and once a year for three years, in case of THA. Weight bearing, anteroposterior radiographs were performed with the patient standing on both legs. The patient's feet were $15^{\circ} \pm 5^{\circ}$ internally rotated. The x-ray beam was horizontal, perpendicular to the table. The source to film distance was 100 cm.

Joint space width measurements

All the films were collected. Before reading, the films of a given patient were placed side by side on a light box. The patient's identity and the date were masked using adhesive tape. The date was replaced by a code. The films were thus assessed blind to the chronology by a single reader. The reader measured JSW in mm at the narrowest point between the distal margin of

the condylar cortex for the femoral surface and the margin of the bright radiodense band of the subchondral cortex in the floor of the articular fossa for the acetabulum, using a 0.1 mm graduated magnifying glass directly laid over the film.

Data analysis

Only one hip was analyzed in each patient. In the study population, JSW measurements and the number of THA were collected at 1, 2 and 3 years. At a given time, patients with missing values were patients with no radiograph available for JSW measurements, patients with THA, or patients lost to follow-up (lack of JSW measurements and information on THA). Hip OA progression was evaluated using five different parameters: 1) JSW measurements in mm, considering patients with THA as having a missing value for JSW (JSW1), 2) JSW measurements in mm, considering patients with THA as having JSW=0 mm (JSW2), 3) number of patients with hip OA progression, defined as JSW changes greater than a given cutoff value or THA performed (JSN+THA), 4) number of patients with hip OA progression, defined as JSW changes greater than a given cut off value, considering patients with THA as having a missing value for JSW (JSN), and 5) time to THA (THA), where data were considered as censored data.

The progression rates for all these parameters were calculated for different follow-up times (1, 2 and 3 years) as mean JSW progression, percentage of patients who progressed, and THA rates. The value of the cut off was 0.5 mm, which defines radiological progression of hip OA (8).

Calculations of sample sizes

We calculated sample sizes supposing a clinical trial evaluating the effect of reducing the progression rate of a given DMOAD in hip OA. We considered a randomized, placebo-

controlled, parallel group design with patients randomized to the drug or placebo groups. We tested treatment effects of 50% and 30% on progression rates between the two groups. We assumed that such treatment effects would be constant regardless of the duration of study and that the percentage of assessable patients would be the same regardless of the study group. We set the study power to 80% with a type I error of 5%. Sample size calculations were determined for different study durations (1, 2 and 3 years). Progression rates observed in our study population were considered as placebo group values. The following parameters were tested for these sample size calculations (15): 1) JSW1, 2) JSW2, 3) JSN+THA, 4) JSN; and 5) THA as survival data at 1, 2, and 3 years.

Determination of person-years follow-up

To estimate the follow-up of patients according to different clinical trial durations (i.e., 1, 2, and 3 years), the number of person-years follow-up was used for each parameter (16). We chose this method because, in the population used, patients experienced various lengths of follow-up. That is, despite a theoretically identical follow-up for all patients included in the clinical trial, the follow-up of some patients would be reduced, in case of radiographs not performed, THA or loss-to-follow-up. Thus, patients with all evaluations counted for the whole study duration considered. Patients with missing values at the end of one year follow-up were considered as non-available for evaluation at that time, but nevertheless counted for half this year. In a 3-year study, patients with missing values at 3 years only counted for 2 and half years. In a 2-year study, patients with missing values at 2 years only counted for 1 and half years. In a 1-year study, patients with missing values at 1 year counted for half a year.

Results

At entry, 252 patients were assessable. At 1, 2, and 3 years, there were 13, 23, and 46 patients, respectively, for whom JSW could not be measured (table 2). Information on THA was not available in 7, 9, and 14 patients (lost to follow-up) at 1, 2, and 3 years, respectively. The number of patients used for assessing hip OA progression varied with the parameter during each study duration (table 3). The percentage of patients assessable every year (that is, at 1, 2, and 3 years) was 60.3% with JSW1, 81.7% with JSW2, 67.1% with JSN, and 88.5% with JSN+THA. For THA, lost to follow-up patients were censored at the date they were lost.

Sample size estimations

Table 4 shows the estimations of sample sizes required per group for three different study durations, using five different parameters. Sample sizes estimated using a 30% treatment effect were about three-fold those estimated using a 50% treatment effect with all parameters. From 1- to 3-year study durations, sample sizes needed would be reduced by 43% with JSW1 (i.e., from 215 to 123 subjects), 57% with JSW2, 61% with JSN+THA, 47% with JSN, and 67% with THA.

Number of person-years follow-up

The number of person-years follow-up with JSW1 and JSW2 increased with increasing study duration (Table 4). This increase was nonetheless greater between 1- and 2-year study durations than between 2- and 3-year study durations. The number of person-years follow-up varied less with JSN+THA and with THA between study durations. With a tested treatment effect of 50%, the differences in the number of person-years follow-up between 1- and 2-year study durations on the one hand and between 1- and 3-year study durations on the other, were

respectively 28.5% and 33.4% with JSW1, 17% and 17.6% with JSW2, 7.7% and 11.7% with JSN+THA, 28.1% and 27.5% with JSN. Such differences were often smaller with a tested treatment effect of 30%. The differences in the number of person-years follow-up between 2- and 3-year study durations were less than 10% with all these parameters.

With THA, the differences in person-years between study durations were less than 10%.

Discussion

We estimated sample sizes needed per group in a clinical trial aiming at assessing a DMOAD in hip OA, supposing a randomized, placebo controlled design, and using 5 different parameters. With a 30% tested treatment effect, between 112 and 1675 patients would be required according to study duration and parameter. Between 2- and 3-year study durations compared to a 1-year study duration, the number of person-years follow-up differed by less than 15% with parameters taking into account missing values (JSW2, JSN+THA, and THA) and more than 20% with JSW1 and JSN. Such differences were less than 10% between 2- and 3-year study durations for all but one parameter (10.9% with JSW1).

In DMOAD clinical trials, the calculation of sample sizes required should be based on realistic estimates of the outcome measures (4,7). Available statistical tables for sample size calculations, and clinical or epidemiological studies that have measured the rate of OA progression may help to determine these estimates (4,7,17,18). However, missing values for primary outcome measures can involve a large part of the patients included in such long-term clinical trials. Statistical methods such as the last observation carried forward, the LOCF method, can be used to overcome this problem in clinical trials of symptomatic drugs. In this method, the missing value for a given patient is replaced by the last value available. In

DMOAD clinical trials, however, such a method may lead, for example, to consider prematurely withdrawn patients as having wider JSW than completers.

In this study, we tested different parameters considering problems raised by missing values when calculating sample sizes in DMOAD clinical trials in hip OA. Using JSW1 did not prevent the loss of information due to THA, radiographs not done, or loss to follow-up. Using JSW2, considering patients with THA as having JSW=0 mm, allowed to partly overcome this problem. JSW2 was nevertheless associated with an increase in variability (table 3). Moreover, considering patients with THA as having JSW=0 mm is a debatable choice. Another choice could be to obtain a radiograph immediately before THA. However, this is not always possible.

JSW can be also expressed as a qualitative variable by defining a cut off permitting to assess radiological progression (8,9,19). Thus, the definition of JSN was based on the reliability of JSW measurements from a wide spectrum of hip OA patients. Moreover, this definition has been demonstrated to correlate with changes in clinical status of the patients and to be related to demographic and specific characteristics of hip OA (8). However, its value may be study-dependent (radiological procedure, measuring instrument, reader, and the sample of studied patients) (9), and results obtained with JSN can be generalized only to the population from which the sample is taken. However, our sample included a large spectrum of patients who are seen by rheumatologists and included in clinical trials.

Although THA and JSN are considered equivalent in the parameter JSN+THA, it seems attractive to use it. The information of patients having THA and partly that of those not x-rayed can be used with this parameter. Furthermore, its use may be relevant, since patients with a change greater than 0.5 mm in JSW or having THA in a 1-year follow-up can be considered as having progressed.

The requirement of THA can depend on concurrent medical and non-medical factors not related to hip OA. However, this composite parameter may combine symptomatic (painful and disabling disease) and structural information (advanced joint space narrowing) (5). Thus, although it is less specific than JSW measurements, this parameter can avoid the bias of considering inconsistent hip OA progression that could occur with JSW measurements. Additionally, it is easier to obtain than JSW measurements and has been shown to be valid as an outcome measure (10). Finally, THA and JSW measurements can be used together in DMOAD clinical trials, separately or combined as with JSN+THA.

Although all these parameters are related to structural changes, they should not be compared to each other on sample size estimations. Such comparisons may be not clinically relevant since these parameters do not measure the same events. For example, testing a 50% treatment effect with JSW measurements would mean reducing the mean progression rate of JSW of the group of patients by 50% whereas with JSW+THA, it would mean reducing the number of patients who progress by 50%. Thus, when planning such clinical trials, investigators should consider not only sample sizes but also the clinical relevance of the parameters used.

Our aim was not to define study durations of DMOAD clinical trials in hip OA. The adequate duration of these clinical trials is currently contentious. The results of this study show that sample size or number of person-years follow-up should be considered for this definition. With JSW1, JSW2 and JSN, the differences in number of person-years follow-up might be negligible between 2- and 3-year study durations whereas those between 1- and 2- or 3-year study durations might be consistent. With THA, such differences might be negligible despite serious differences in sample size estimates between study durations. Actually, study duration should depend also on the mode of action of the drug, the anticipated response rate, the primary outcome, and the length of time needed to show a difference in comparison with the control group (5,20). As an example, a 1-year duration can be advocated by some

investigators as suitable for evaluating DMOAD. However, most of the progression observed in patients having rapid radiographic progression may occur during the first year of follow-up. In such patients, DMOAD might be ineffective.

In this study, we have shown that investigators should appropriately manage the problem raised by missing values when designing DMOAD clinical trials. Our results can help them in the planning stage of these clinical trials.

The authors thank Dr Laurent Berdah, Pr Bernard Mazières, Dr Minh Nguyen, Pr Eric Vignon, members of the steering committee of the study from which patients used were taken.

References:

1. Kelsey JL, Hochberg MC. Epidemiology of chronic musculoskeletal disorders. *Annu Rev Public Health* 1988; 9: 379-401.
2. Pavelká K, Gatterova J, Gollerova V, Urbanova Z, Sedlackova M, Altman RD. A 5-year randomized controlled, double-blind study of glycosaminoglycan polysulphuric acid complex (Rumalon) as a structure modifying therapy in osteoarthritis of the hip and knee. *Osteoarthritis Cartilage* 2000; 8: 335-42.
3. Altman RD, Howell DS. Disease-modifying osteoarthritis drugs. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998: 417-28.
4. Altman R, Brandt K, Hochberg M, Moskowitz R. Recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis: Design and conduct of clinical trials in patients with osteoarthritis*. *Osteoarthritis Cartilage* 1996; 4: 217-243.
5. Bellamy N, Kirwan J, Boers J, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for the future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT III. *J Rheumatol* 1997; 24: 799-802.
6. Group for the Respect of Ethics and Excellence in Science (GREES): Osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis* 1996; 55: 552-557.
7. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes CJ, Pelletier J-P. Guidelines for testing slow acting drugs in osteoarthritis. *J Rheumatol* 1994; (suppl 41) 21: 65-73.
8. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, Vignon E. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996; 55: 356-62.

9. Ravaud Ph, Giraudeau B., Auleley G-R, Drape J-L, Rousselin B, Paolozzi L, Chastang CI, Dougados M. Variability in knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis* 1998; 57: 624-29.
10. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, Vignon E. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? *J Rheumatol* 1999; 26: 855-61.
11. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification et reporting of the hip. *Athritis Rheum* 1991;34:505-14.
12. Altman RD. The classification of osteoarthritis. *J Rheumatol* 1995; (suppl 53) 22: 42-3.
13. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. *Scand J Rheumatol* 1987; suppl 65: 85-89.
14. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957; 16: 494-502.
15. Machin D, Campbell MJ, Fayers PM, Pinol APY. Sample size tables for clinical studies. Sd ed., Oxford: Blackwell, 1997.
16. Armitage P, Berry G. Statistical methods in medical research. 3 rd ed., Oxford: Blackwell, 1994.
17. Bellamy N. Design and of clinical trials for evaluation of DMOADs and of new agents for symptomatic treatment of osteoarthritis. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998: 531-42.
18. Bellamy N. Standard deviation tables for sample size calculation in rheumatoid arthritis, osteoarthritis, ankylosing spondylitis and fibromyalgia clinical trials. In: Bellamy N, ed. *Musculoskeletal clinical metrology*. Dordrecht: Kluwer Academic Publishers, 1993: 337-

19. Ravaud P, Giraudeau B, Auleley G-R, Edouard-Noël R, Dougados M, Chastang Cl. Assessing smallest detectable change over time in continuous structural outcome measures: application to radiological change in knee osteoarthritis. *J Clin Epidemiol* 1999; 52: 1225-30.
20. Mazzuca S, Brandt KD. Plain radiography as an outcome measure in clinical trials involving patients with knee osteoarthritis. *Rheum Dis Clin North Am* 1999; 25: 467-80.

Table 1: Baseline clinical and radiological characteristics of hip osteoarthritis patients. Values are either mean±SD or n (%).

Age (years)	62.7 ± 7.1
Sex (m/f)	104/148
Body mass index (kg/m ²)	25.6 ± 3.6
Pain (VAS 0-100)	45.6 ± 19.8
Lequesne functional index (0-24)	7.8 ± 2.7
Kellgren and Lawrence grading system	
Grade 1 (%)	4 (1.6)
Grade 2 (%)	175 (69.4)
Grade 3 (%)	72 (28.6)
Grade 4 (%)	1 (0.4)

Table 2: Main reasons and number of patients not assessable for joint space width (JSW) measurements in hip osteoarthritis in 252 patients.

Time of measurements	Hip arthroplasty	Radiograph not available	Lost to follow-up	Total
At year 1	20	6	7	33
At year 2	40	14	9	63
At year 3	54	32	14	100

Table 3: Hip osteoarthritis progression assessed with five different parameters in 252 patients.

Study duration	JSW1	JSW2	Parameters* JSN+THA	JSN	THA
One year					
Assessable patients, n	219 (86.9%)	239 (94.8%)	239 (94.8)	219 (86.9%)	245 (97.2%)
Progression rate	-0.33±0.57	-0.48±0.78	84 (35.1%)	64 (29.2%)	20 (8.2%)
Two years					
Assessable patients, n	189 (75%)	229 (90.9%)	223 (92.5%)	193 (76.6)	243 (96.4%)
Progression rate	-0.45±0.62	-0.73±0.92	117 (52.5%)	77 (39.9%)	40 (16.5%)
Three years					
Assessable patients, n	152 (60.3%)	206 (81.7%)	223 (88.5%)	169 (67.1%)	238 (94.4%)
Progression rate	-0.62±0.67	-0.98±0.96	146 (65.5%)	92 (54.4%)	54 (22.7%)

* JSW1= Joint space width (JSW) measurements in mm considering patients with total hip arthroplasty (THA) as having missing values for JSW; JSW2= JSW measurements in mm considering patients with THA as having JSW=0 mm; JSN+THA = number of patients with hip progression defined as joint space narrowing greater than 0.5 mm (JSN>0.5 mm) or THA performed; JSN= number of patients with hip progression defined as JSN> 0.5 mm considering patients with THA as having missing value for JSW; THA= number of patients having THA at 1, 2 and 3 years.

Table 4: Estimated sample sizes (number of person-years) per group for 1-, 2-, and 3-year durations of study, with 30% and 50% treatment effects in progression rates between two groups in hip osteoarthritis clinical trials, using five different parameters.

	Parameters*				
	JSW1	JSW2	JSN+THA	JSN	THA
Effect of 50% for					
1 year	215 (200.5)	178 (173)	105 (102)	144 (134.5)	559 (551)
2 years	160 (280.5)	109 (208.5)	58 (110.5)	108 (187)	261 (510)
3 years	123 (300.5)	76 (210)	41 (115.5)	76 (185.5)	185 (537.5)
Effect of 30% for					
1 year	589 (550)	516 (502.5)	314 (305.5)	446 (416.5)	1675 (1651.5)
2 years	412 (721.5)	304 (578)	168 (320.5)	319 (518.5)	759 (1539.5)
3 years	327 (809.5)	212 (586.5)	112 (314.5)	218 (540)	563 (1636)

* JSW1= Joint space width (JSW) measurements in mm considering patients with total hip arthroplasty (THA) as having missing values for JSW; JSW2= JSW measurements in mm considering patients with THA as having JSW=0 mm; JSN+THA = number of patients with hip progression defined as joint space narrowing greater than 0.5 mm (JSN>0.5 mm) or requirement for THA; JSN= number of patients with hip progression defined as JSN> 0.5 mm considering patients with THA as having missing value for JSW; THA= number of patients requiring THA at 1, 2, and 3 years.

CHAPITRE VI

CONCLUSION

Cette thèse avait pour ambitions de contribuer à optimiser la mesure de la hauteur de l'interligne articulaire dans la coxarthrose.

L'étude de la concordance entre les mesures de la hauteur de l'interligne articulaire sur les radiographies réalisées en position debout et couchée nous a permis de montrer qu'il n'existait pas de surévaluation des mesures sur les radiographies réalisées dans une position par rapport à l'autre. Par ailleurs, nous avons pu estimer l'erreur de mesure sur la hauteur de l'interligne articulaire entre ces deux positions.

Nous avons ensuite montré que la modification du centrage du rayon incident ou de la rotation interne des membres inférieurs du patient pendant la réalisation des radiographies standards du bassin s'accompagnait d'une augmentation de la variabilité des mesures. Nous avons pu ainsi proposer des recommandations permettant d'améliorer la reproductibilité de cet examen. Enfin, en étudiant l'intérêt de standardiser les conditions d'acquisition des radiographies et de la mesure, nous avons pu proposer un seuil de progression de la coxarthrose.

Nous avons également comparé plusieurs méthodes de lecture des radiographies (radiographies successives d'un même patient lues par couples chronologiquement ordonnées, par couples en aveugle de la chronologie, radiographies lues isolément) pour évaluer la progression de la coxarthrose (score de Kellgren et Lawrence, pincement articulaire, mesure de l'interligne). Ainsi, nous avons pu montrer que la méthode de lecture avait une influence importante sur les critères qualitatifs et que la connaissance de la chronologie des radiographies d'un même patient n'apportait pas d'information importante pour évaluer la hauteur de l'interligne articulaire.

Par ailleurs, nous avons comparé l'utilisation associée d'une méthode lecture des radiographies avec le marquage du site de mesure sur l'interligne pour évaluer la progression de la hauteur l'interligne articulaire. Nous avons montré l'intérêt des radiographies lues par couples, en aveugle de la chronologie, associée au marquage du site de mesure pour évaluer la progression de l'interligne articulaire dans les études longitudinales.

Enfin, nous avons estimé la taille des essais thérapeutiques visant à évaluer l'efficacité des traitements anti-arthrosiques à visée structurale en utilisant différents critères de jugement dans la coxarthrose.

Les résultats de ces travaux sont déjà appliqués dans les essais thérapeutiques concernant la coxarthrose, qu'il s'agisse de la standardisation de la position du patient, de la procédure technique d'acquisition des radiographies ou qu'ils s'agisse de la lecture des radiographie.

La méthodologie utilisée dans ces travaux a été antérieurement utilisée au genou. Nous confirmons son intérêt dans l'évaluation de la mesure.

D'autres applications médicales posent les mêmes problèmes que ceux étudiés dans ces travaux. Elles pourraient être évaluées avec cette méthodologie.

Bibliographie

Adams ME, Wallace CJ. Quantitative imaging of osteoarthritis. *Semin Arthritis Rheum* 1991 ; 20 (Suppl 2) : 26-39.

Ahlback S. Osteoarthritis of the knee: a radiographic investigation. *Acta Radiol* 1968 ; 277 (suppl) : 7-72.

Altman R, Fries JF, Bloch DA, et al. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987 ; 30 : 1214-1225.

Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum* 1986 ; 29 : 1039-1049.

Altman R, Alarcon G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34: 505-514.

Altman R, Hochberg M, Murphy W, et al. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995 ; 3 (Suppl A) : 3-71.

Altman R, Brandt T, Hochberg M, et al. Design and conduct of clinical trials of patients with osteoarthritis. Recommendations from a task force of the Osteoarthritis Research Society. *Osteoarthritis Cartilage* 1996; 4: 217-243.

Altman RD, Howell DS. Disease-modifying osteoarthritis drugs. In: Brandt KD, Doherty M, Lohmander LS, eds. Osteoarthritis. New York: *Oxford University Press*, 1998: 417-428.

Armitage P, Berry G. Statistical methods in medical research. *Blackwell Scientific Publications*, Third edition, Oxford, 1994.

Ayral X, Dougados M, Listrat V, et al. Chondroscopy: a new method for scoring chondropathy. *Semin Arthritis Rheumatism* 1993 ; 22 : 289-297.

Ayral X, Dougados M, Listrat V, et al. Arthroscopic evaluation of chondropathy in osteoarthritis of the knee. *J Rheumatol* 1996 ; 23 : 698-706.

Ayral X, Ravaud P. Evolution de l'arthrose: méthodes d'évaluation objective par l'imagerie médicale et par l'arthroscopie. Application à la gonarthrose fémoro-tibiale interne. *La Presse Médicale* 1998; 27: 1491-1498.

Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep* 1966 ; 19 : 3-11.

Bellamy N, Carette S, Ford PM, et al. Osteoarthritis antirheumatic drugs trials: I- Effects of standardization procedures on observer dependent outcome measures. *J Rheumatol* 1992; 19: 436-443.

Bellamy N. Musculoskeletal Clinical Metrology. *Kluwer Academic Publishers*, 1993.

Bellamy N. Design of clinical trials for evaluation of DMOADs and new agents for symptomatic treatment of osteoarthritis. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998: 417-428.

Bernard PM, Lapointe C. Mesures statistiques en épidémiologie. *Presses de l'Université du Québec*, Québec, 1995.

Blackburn WD, Bernreuter WK, Rominger M, et al. Arthroscopic evaluation of knee articular cartilage: a comparison with plain radiographs and magnetic resonance imaging. *J Rheumatol* 1994 ; 21 : 675-679.

Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986 ; i : 307-310.

Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med* 1990 ; 20 : 337-340.

Bland JM, Altman DG. Comparing two methods of clinical measurement: a personal history. *Int J Epidemiol* 1995; 24 (suppl ,1): S7-S14.

Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Meth Med Res* 1999; 8: 135-160.

Brandt KD, Fife RS, Braunstein EM, Katz B. Radiographic grading of the severity of knee osteoarthritis: relation of Kellgren and Lawrence grade to a grade based on joint space narrowing and correlation with arthroscopic evidence of articular degeneration. *Arthritis Rheum* 1991 ; 34 : 1381-1386.

Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992 ; 304 : 1491-1494.

Broderick S, Turner DA, Renfrew DL, et al. Severity of articular cartilage abnormality in patients with osteoarthritis: evaluation with fast spin echo MR versus Arthroscopy. *AJR* 1994 ; 162 : 99-103.

Buckland-Wright JC, Bradshaw CR. Clinical application of high definition microfocal radiography. *Br J Radiol* 1989; 62: 209-217.

Buckland-Wright JC. Quantitative radiography of osteoarthritis. *Ann Rheum Dis* 1994 ; 63 : 268-275.

Buckland-Wright JC, MacFarlane, Lynch JA, et al. Joint space width measures cartilage thickness in osteoarthritis of the knee: high resolution plain film and double contrast macroradiographic investigation. *Ann Rheum Dis* 1995a ; 54 : 263-268.

Buckland-Wright JC, McFarlane DG, Williams SA, et al. Accuracy and precision of joint space width measurement in standard and macro radiographs of osteoarthritis knees. *Ann Rheum Dis* 1995b ; 54 : 872-880.

Buckland-Wright JC. Quantitative radiography: Osteoarthritis-microfocal. *Baillière's Clinical Rheumatology* 1996; 10: 415-420.

Buckland-Wright JC. Quantitative of radiographic changes. In: Brandt KD, Doherty M, Lohmander LS, eds. Osteoarthritis. New York: *Oxford University Press*, 1998: 459-472.

Buckland-Wright JC. Radiographic assessment of osteoarthritis: Comparison between existing methodologies. *Osteoarthritis Cartilage* 1999; 7: 430-433.

Burdock EI, Fleiss JL, Hardesty AS. A new view of inter-observer agreement. *Pers Psychol* 1963 ; 16 : 373-384.

Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993 ; 46 : 423-424.

Cagriota-Scanderberg A, De Micheli V, Scarabe MG, et al. Precision of sonographic measurement of articular cartilage: inter and intraobserver analysis. *Skeletal Radiol* 1996; 25: 545-549.

Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. *J Clin Epidemiol* 1990 ; 43 : 551-558.

Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960 ; 20 : 37-46.

Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968 ; 70 : 213-20.

Conrozier T, Balblanc JC, Tron AM, et al. Détermination de la surface et de l'épaisseur articulaires coxo-femorales par un analyseur automatique d'images numérisées. *Rev Rheum* 1990;58:695.

Conrozier T, Tron AM, Mathieu P, Vignon E. Quantitative assessment of radiographic normal and osteoarthritic hip joint space. *Osteoarthritis Cartilage* 1995 ; 3 (suppl A) : 81-87.

Conrozier T, Tron A-M, Vignon E. Coxarthroses. *Rev Prat* 1996; 46: 2201-2205.

Conrozier T, Lequesne MG, Tron AM, et al. The effects of position on the radiographic joint space in osteoarthritis of the hip. *Osteoarthritis Cartilage* 1997 ; 5: 17-22.

Conrozier T, Jousseau CA, Mathieu P, et al. Quantitative measurement of joint space narrowing progression in hip osteoarthritis: a longitudinal retrospective study of patients treated by total hip arthroplasty. *Br J Rheumatol* 1998; 37: 961-968.

Croft P, Cooper C, Wickham C, et al. Defining osteoarthritis of the hip for epidemiologic studies. *Am J Epidemiol* 1990 ; 132 : 514-522.

Dacre JE, Coppock JS, Herbert KE, et al. Development of a new radiographic scoring system using digital image analysis. *Ann Rheum Dis* 1989a; 48 : 194-200.

Dacre JE, Huskisson EC. The automatic assessment of knee radiographs using digital image analysis. *Br J Rheumatol* 1989b ; 28 : 506-510.

Dieppe PA, Cushnaghan J, Jasani MK, et al. A two year, placebo-controlled trial on non steroidal anti-inflammatory therapy in osteoarthritis of the knee joint. *Br J Rheumatol* 1993; 32 : 595-600.

Dieppe P, Altman RD, Buckwalter JA, et al. Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. In Kuettner KE, Golberg VM, eds. Osteoarthritis disorders. Rosemont, II: American Academy of Orthopaedic Surgeons, 1994: 481-496.

Dieppe P, Altman RD, Buckwalter JA, et al. Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. New Horizons in osteoarthritis, Kuettner K and Goldberg V, eds. *American Academy of orthopaedic Surgeons, Rosemont, USA* 1994 ; 481-496.

Dieppe PA. Recommended methodology for assessing the progression of the osteoarthritis of the hip and knee joint. *Osteoarthritis Cartilage* 1995a ; 3 : 73-77.

Dieppe P, Brandt KD, Lohmander S, et al. Detecting and measuring disease modification in osteoarthritis. The need for standardized methodology. *J Rheumatol* 1995b ; 22 : 201-203.

Disler DG, Mc Cauley TR, Wirth CR, et al. Detection of knee hyaline articular cartilage defects using fat-suppressed three-dimensional spoiled gradient-echo MR imaging: comparison with standard MR imaging and correlation with arthroscopy. *AJR* 1995 ; 165: 377-382.

Dougados M, Listrat V, Nguyen M, et al. Analyses de la surface de l'interligne articulaire après digitalisation des clichés radiographiques au cours de la gonarthrose fémoro-tibiale. Actualités en physiologie et pharmacologie articulaires. A Gaucher 1991 ; *Éditions Masson* 286 91.

Dougados M, Gueguen A, Nguyen M, et al. Longitudinal radiologic evaluation of osteoarthritis of the knee. *J Rheumatol* 1992 ; 19 : 378-384.

Dougados M, Ayrat X, Listrat V, et al. The SFA for assessing articular cartilage lesions at arthroscopy of the knee. *Arthroscopy* 1994, 10 : 69-77.

Dougados M, Villers Ch, Amor B. Sensibilité au changement de différentes techniques de mesure de la sévérité radiologique au cours de la coxarthrose. *Rev Rhum [Ed Fr]* 1995; 62: 179-184.

Dougados M, Gueguen A, Nguyen M, et al. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996a; 55: 356-362.

Dougados M. La mesure. Initiation du clinicien au langage statistique et méthodologique. *Expansion Scientifique Française*, Paris, 1996b.

Dougados M, Gueguen A, Nguyen M, et al. Requirement for total hip arthroplasty: An outcome measure of hip osteoarthritis? *J Rheumatol* 1999; 26: 855-861.

Drape J-L, Pessis E, Auleley G-R, et al. Quantitative MR imaging: Evaluation of chondropathy in osteoarthritic knees. *Radiology* 1998; 208: 49-55.

Ebel RL. Estimation of the reliability of ratings. *Psychometrika* 1951 ; 16 : 407-424.

Evison G, Reilly PA, Gray J, et al. A comparison of erect and supine radiographs of the hip. *Br J Rheumatol* 1987; 26: 393-394.

Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: *WB Saunders* 1985.

Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990 ; 43 : 543-49.

Felson DT. Epidemiology of hip and knee osteoarthritis. *Epidemiol Rev* 1988; 10: 1-18.

Felson DT, Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 1988; 41: 1343-1355.

Fermanian J. Mesure de l'accord entre deux juges: cas qualitatif. *Rev Epidem et Sant» publ* 1984a ; 32 : 140-147.

Fermanian J. Mesure de l'accord entre deux juges: cas quantitatif. *Rev Epidem et Sant» publ* 1984b ; 32 : 408-413.

Fernandez-Madrid F, Karvonen RL, Teitge RA, et al. MRI features of osteoarthritis of the knee. *Magnetic Resonance Imaging* 1994 ; 12 : 703-709.

Ferrara R, Priolo F, Cammisa M, et al. Clinical trails in rheumatoid arthritis: Methodological suggestions for assessing radiographs arising from the GRISAR study. *Ann Rheum Dis* 1997; 56: 608-612.

Fife RS, Brandt KD, Braunstein EM, et al. Relationship between arthroscopic evidence of cartilage damage and radiographic evidence of joint space narrowing in early osteoarthritis of the knee. *Arthritis Rheum* 1991 ; 34 ; 4 : 377-382.

Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973 ; 33 : 613-619.

Fleiss JL. Statistical Methods for Rates and Proportions, second edition. New York : *Wiley* 1982.

Fox KM, Hochberg MC, Resnick CS, et al. Severity of radiographic findings in hip osteoarthritis associated with total hip arthroplasty. *J Rheumatol* 1996; 23: 693-697.

Glimet T, Mass» JP. Variation intra-observateur de la mesure de l'interligne f»moro-tibial interne dans la gonarthrose. *Rev Rheum* 1989 ; 56 : 715-716.

Group for the respect of ethics and excellence in science (GREES): osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis* 1996 ; 55 : 552-557.

Handelberg F, Shahabpour M, Casteleyn PP. Chondral lesions of the patella evaluated with computed tomography magnetic resonance imaging and arthroscopy. *Arthroscopy* 1990 ; 6 : 24-29.

Hansson G, Jerre R, Sanders SM, et al. Radiographic assessment of coxarthrosis following slipped capital femoral epiphysis. A 32-year follow-up study of 151 hips. *Acta Radiol* 1993; 34: 117-123.

Hart DJ, Spector TD. Radiographic criteria for epidemiologic studies of osteoarthritis. *J Rheumatol* 1995; (suppl 43) 22: 46-48.

Hemodsson J. Roentgen appearance of coxarthrosis. *Acta Orthop Scandinav* 1970; 41: 169-187.

Hochberg MC, Altman RD, Brandt KD, et al. Design and conduct of clinical trials in osteoarthritis: Preliminary recommendations from a task force of the Osteoarthritis Research Society. *J Rheumatol* 1997; 24: 792-794.

Holder J, Trudell D, Pathria MN, et al. Width of articular cartilage of the hip: Quantification by using fat-suppression spin-echo MR imaging in cadavers. *AJR* 1992; 159: 351-355.

Hutton CW, Vennart W. Osteoarthritis and magnetic resonance imaging: potential and problems. *Ann Rheum Dis* 1995 ; 54 : 237-243.

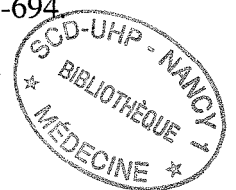
Jewell FM, Watt I, Doherty M. Plain radiographic features of osteoarthritis. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: Oxford University Press, 1998: 217-237.

Jonsson K, Buckland K, Helvie M, et al. Precision of hyaline cartilage thickness measurements. *Acta Radiol* 1992 ; 33 : 234-239.

Jorring K. Osteoarthritis of the hip. Epidemiology and clinical role. *Acta Orthop Scand* 1980; 51: 523-530.

Kallman DA, Wigley FM, Scott WW Jr, et al. New radiographic grading scales for osteoarthritis of the hand. Reliability for determining prevalence and progression. *Arthritis Rheum* 1989; 32: 1584-1591.

Katz JN, Guadagnoli E, Liang M, et al. Differences between men and women undergoing major orthopaedic surgery for degenerative arthritis. *Arthritis Rheum* 1994; 37: 687-694.



- Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. *Ann Rheum Dis* 1957 ; 16 : 494-502.
- Kelsey JL, Hochberg MC. Epidemiology of chronic musculoskeletal disorders. *Ann Rev Public Health* 1988; 9: 379-401.
- Kirwan JR, Cushnagan J, Dacre J, et al. Progression of joint space narrowing in knee osteoarthritis. *Arthritis Rheum* 1992 ; 35 : S134 (Abstract).
- Kramer MS, Feinstein AR. The biostatistics of concordance. *Clin Pharmacol Ther* 1981 ; 29: 111-123.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977 ; 33 : 159-174.
- Lane NE, Nevitt MC, Genant HK, et al. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. *J Rheumatol* 1993 ; 20: 1911-1918.
- Leach RE, Greff T, Sibert FJ. Weight-bearing radiography in osteoarthritis of the knee. *Radiology* 1970 ; 97 : 265-268.
- Leclerc A, Papoz L, Bréart G, et al. Dictionnaire d'Epidémiologie. *Frison-Roche*, Paris, 1990.
- Ledingham J, Dawson S, Preston B, et al. Radiographic progression of hospital referred osteoarthritis of the hip. *Ann Rheum Dis* 1993; 52: 263-267.
- Lee J, Koh D, Ong CN. Statistical evaluation of agreement between two methods for measuring a quantitative variable. *Comput Biol Med* 1989 ; 19: 61-70.
- Lee J. Evaluating agreement between two methods for measuring the same quantity: a response. *Comput Biol Med* 1992 ; 22: 369-371.
- Lequesne M, Wilhelm F. Méthodologie pour le clinicien. Précis et glossaire. *EUULAR*, Bâle, 1989.
- Lequesne M, Brandt K, Bellamy N, et al. Guidelines for testing slow-acting and disease modifying drugs in osteoarthritis. *J Rheumatol* 1994a; (suppl 41) 21 : 65-73.

Lequesne M. Quantitative measurements of joint space during progression of osteoarthritis : "chondrometry". In Kuettner KE, Goldberg VM, ed. Osteoarthritic disorders. Symposium of the American Academy of orthopaedic surgeons, Monterey, CA, 1994. Rosemont, IL : *American Academy of orthopaedic surgeons Publications*, 1994b; 427-444.

Lequesne M. La chondrométrie. Evaluation quantitative de l'épaisseur de l'interligne et de sa dégradation annuelle dans la coxarthrose. *Rev Rhum [Ed Fr]* 1995; 62: 165-168.

Lequesne M, Cadet C, Auleley G-R. Mesure à la loupe du pincement de l'interligne articulaire (PI) dans la coxarthrose (Cox). Amélioration de la reproductibilité. *Rev Rhum [Ed Fr]* 1999; 66: 637-(74) (Abstract).

Li KC, Higgs J, Aisen AM, et al. MRI in osteoarthritis of the hip: Gradations of severity. *Magn Reson Imaging* 1988; 6: 229-236.

Lynch JA, Buckland-Wright JC, MacFarlane DG. Precision of joint space with measurement in knee osteoarthritis from digital image analysis of high definition macroradiographs. *Osteoarthritis Cart* 1993 ; 1 : 209-218.

Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; 126 : 161-169.

Massé JP, Glimet T, Nguyen NJ et al. Abstract: Précision de la mesure de l'interligne dans la gonarthrose fémoro-tibiale interne: Reproductibilité. *Rev Rhum* 1990 ; 57: 680 (Abstract).

Mazières B, Levour'ch P, Boichut D. Abstract: Mesure de l'interligne du genou arthrosique. Reproductibilité. Coefficients de variation (Cv) intra et inter-observateur. *Rev Rhum* 1990a ; 57: 680 (Abstract).

Mazières B, Levour'ch P, Bunouf P. Mesures des interlignes du genou arthrosique. Variations en un an. *Rev Rhum* 1990b ; 57: 680 (Abstract).

Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med* 1994; 13: 2465-2476.

Murphy WA, Alman RD. Updated osteoarthritis reference standard. *J Rheumatol* 1995; (suppl 43) 22: 56-59.

Oliveira SA, Felson DT, Reed JI, et al. Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis Rheum* 1995; 38: 1134-1141.

Papoz L. Qualité des critères de jugement. In Bouvenot G, Vray M. Essais cliniques: Théorie, pratique et critique. *Médecine-Sciences Flammarion*, Paris, 1994: 25-38.

Peterfy CG, van Dijke CF, Janzen DL, et al. Quantification of articular cartilage in the knee by pulsed saturation transfer and fat-suppressed MRI: optimization and validation. *Radiology* 1994; 192: 485-491.

Peterfy CG, Majumdar S, Lang P, et al. MR imaging of the arthritic knee : improved discrimination of cartilage, synovium and effusion with pulsed saturation transfer and fat-suppressed T1-weighted sequences. *Radiology* 1994 ;191: 413-419.

Peterfy CG, Genant HK. Emerging applications of magnetic resonance imaging for evaluating the articular cartilage. *Radiol Clin N Am* 1996; 34: 195-213.

Peterfy CG. Magnetic resonance imaging. In: Brandt KD, Doherty M, Lohmander LS, eds. *Osteoarthritis*. New York: *Oxford University Press*, 1998: 473-494.

Petersson IF. Occurrence of osteoarthritis of the peripheral joints in European populations. *Ann Rheum Dis* 1996; 55: 659-664.

Pilch L, Stewart C, Gordon S, et al. Assessment of cartilage volume in the femorotibial joint with magnetic resonance imaging and 3D computer reconstruction. *J Rheumatol* 1994; 21 : 2307-2319.

Preidler KW, Resnick D. Imaging of osteoarthritis. *Radiol Clin North Am* 1996; 34: 259-271.

Ravaud P, Auleley G-R, Amor B, et al. Radiographic assessment of progression in knee osteoarthritis. *Rheumatology in Europe* 1995; suppl 2: 129-131.

Ravaud P, Auleley G-R, Chastang C, et al. Knee joint space width measurement: An experimental study of the influence of radiographic procedure and joint positioning. *Br J Rheumatol* 1996a; 35: 761-766.

Ravaud P, Chastang C, Auleley G-R, et al. Assessment of joint space width in patients with osteoarthritis of the knee: A comparison of four measuring instruments. *J Rheumatol* 1996b; 23: 1749-1755.

Ravaud P, Giraudeau B, Auleley G-R, et al. Radiographic assessment of knee osteoarthritis : Reproducibility and sensitivity to change. *J Rheumatol* 1996c; 23: 1756-1764.

Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. *J Rheumatol* 1997; 24: 786-791.

Ravaud P, Giraudeau B, Auleley G-R, et al. Variability in knee osteoarthritis: Implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis* 1998; 57: 624-629.

Recht MP, Kramer J, Marcelis S, et al. Abnormalities of articular cartilage in the knee: analysis of available MR techniques. *Radiology* 1993; 187: 473-478.

Recht MP, Pirraino DW, Paletta GA, et al. Accuracy of fat-suppressed three-dimension! spoiled gradient-echo FLACH MR imaging in the detection of patellofemoral articular cartilage abnormalities. *Radiology* 1996; 198: 209-212.

Resnick D, Niwayama G. Degenerative diseases of extra-spinal locations. In Resnick D, Niwayame G, eds. *Diagnosis of bone and joint disorders*, second edition. Philadelphia: *Saunders* 1988: 1365-1479.

Sada PN, Rajan P, Jeyaseelan L, et al. Standards for ultrasonographic measurements of the hip joint in Indian adults. *Skeletal Radiol* 1994; 23: 111-112.

Selby K, Peterfy CG, Cohen ZA, et al. In vivo MR quantification of articular cartilage water content : a potential early indicator of osteoarthritis. ed. *Book of Abstracts: Society of Magnetic Resonance* 1995. Berkeley: Society of Magnetic Resonance, 1995 : 204.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428.

Siu D, Cooke TD, Broekhoven CD, et al. A standardized technique for lower links radiography practice, applications and error analysis. *Invest Radiol* 1991; 26: 71-77.

Spector TD, Dacre JE, Harris PA, et al. Radiological progression of osteoarthritis: an 11 year follow up study of the knee. *Ann Rheum Dis* 1992 ; 51 : 1107-1110.

Spector TD, Cooper C. Radiological assessment of osteoarthritis in population studies: Wither Kellgren and Lawrence ? *Osteoarthritis Cartilage* 1994 ; 1 : 203-206.

Spector TD, Hochberg MC. Methodological problems in the epidemiological study of osteoarthritis. *Ann Rheum Dis* 1994; 53: 143-146.

Streiner DL, Norman GR. *Health measurement scales*. Second edition. Oxford University Press ; 1995.

Sui D, Cooke TDV, Broenkhoven LD, et al. A standardized technique for lower limb radiography, practice, applications and error analysis. *Invest radiol* 1991; 26: 71-77.

Tammemagi M, Franck JW, Streiner D. Assessing reproductibility of enzyme-linched immunosorbantamong serologic tests. *J Clin Epidemiol* 1996; 49: 1075-1076.

Tepper S, Hochberg MC. Factors associated with hip osteoarthritis: dat from the first National Health and Nutrition Examination Survey (NHANES-I). *Am J Epidemiol* 1993; 137: 1081-1088.

Teshima R, Otsuka T, Yamamoto K. Effects of nonweight bearing on the hip. *Clin Orthop* 1992; 279: 149-156.

Theiler R, Stucki G, Schutz R, et al. Parametric and non parametric measures in the assessment of the knee and hip osteoarthritis: Interobserver reliability and correlation with radiology. *Osteoarthritis Cartilage* 1996; 4: 35-42.

Thomas RH, Resnick D, Alazraki NP, et al. Compartmental Evaluation of osteoarthritis of the knee. *Radiology* 1975; 116: 585-94.

Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol* 1982; 9: 758-762.

Turula KB, Friberg O, Haajanen J, et al. Weight-bearing radiography in total hip replacement. *Skeletal Radiol* 1985; 14: 200-204.

Tyrrel R, Gluckert K, Pathria M, et al. Fast three dimensional MR Imaging of the knee : comparison with arthroscopy. *Radiology* 1988; 166: 865-72.

Villar RN. Arthroscopy. *BMJ* 1994; 308: 51-53.

Wood PHN. Osteoarthritis in the community. *Clin Rheum Dis* 1976 ; 2 : 495-507.

Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone J Surg [Br]* 1992; 74-B: 287-291.

Nom : Mr AULELEY
Prénom : Guy Robert

DOCTORAT DE L'UNIVERSITÉ HENRI POINCARÉ, NANCY 1

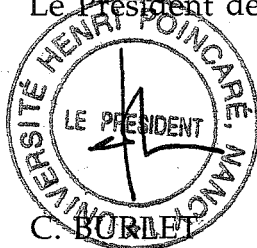
en BIOLOGIE SANTÉ ENVIRONNEMENT

Spécialité : ÉPIDÉMIOLOGIE ET SANTÉ PUBLIQUE

VU, APPROUVÉ ET PERMIS D'IMPRIMER

Nancy, le 13 février 2001 n° 434

Le Président de l'Université



NOM-PRENOM : AULELEY Guy-Robert

DOCTORAT DE L'UNIVERSITÉ HENRI POINCARÉ , NANCY I-ANNÉE 2001

SUJET DE LA THÈSE : CONTRIBUTION A L'OPTIMISATION D'UN CRITÈRE DE JUGEMENT :MESURE DE L'INTERLIGNE ARTICULAIRE DANS LA COXARTHROSE.

MOTS CLÉS : COXARTHROSE, RADIOGRAPHIE STANDARD, INTERLIGNE ARTICULAIRE, REPRODUCTIBILITÉ, STANDARDISATION

RESUMÉ : La mesure de la hauteur de l'interligne articulaire sur les radiographies standards du bassin est le principal critère de jugement pour évaluer la progression de l'arthrose. Ce travail a pour but d'évaluer les différentes sources de variabilité de cette mesure et d'optimiser ses conditions d'utilisation dans la coxarthrose.

Dans une première étude, nous avons étudié l'influence de la mise en charge du patient pendant la réalisation des radiographies sur les mesures de la hauteur de l'interligne articulaire. Nous avons montré qu'il n'avait pas de biais entre les mesures de l'interligne articulaire sur radiographies standards du bassin réalisées en position couchée et debout.

Dans une deuxième étude, nous avons évalué l'influence de la rotation des membres inférieurs et du centrage du rayon incident pendant la réalisation des radiographies du bassin, sur les mesures de la hauteur de l'interligne articulaire. Nous avons montré que la modification de chacun de ces deux paramètres augmentait la variabilité des mesures de l'interligne articulaire. Nous avons aussi évalué l'effet de la standardisation de la procédure d'acquisition des radiographies et de la mesure de l'interligne dans les études prospectives. Nous avons pu proposer des recommandations tendant à améliorer cette mesure. Nous avons également proposé un seuil de progression radiologique de la coxarthrose de 0,50 mm basé sur la variabilité des mesures de la hauteur de l'interligne articulaire sur des radiographies standardisées.

Dans une troisième étude, nous avons étudié l'influence des techniques de lecture des radiographies dans les études longitudinales (ordre chronologique des radiographies, marquage du site de mesure). Nous avons montré qu'il faudrait 10% moins de sujets à inclure dans ces études si les mesures de la hauteur de l'interligne articulaire étaient faites avec marquage du site de mesure, sur des radiographies groupées par patient et lues en aveugle de la chronologie plutôt que lues chacune isolément (en aveugle de l'identité du sujet et de la chronologie des radiographies).

Enfin, en utilisant les différents critères de jugement actuellement proposés, nous avons estimé le nombre de sujets nécessaires à inclure dans les essais thérapeutiques évaluant les traitements anti-arthrosiques à visée structurale dans la coxarthrose. Pour cela, nous avons pris en compte les données manquantes. Nous avons montré que la variation du nombre de personnes-années de suivi était directement liée à la méthode de prise en compte des données manquantes.

JURY: Monsieur le Professeur Serge BRIANÇON (Nancy)

Madame le Professeur Isabelle CHARY-VALCKENAERE (Nancy)

Monsieur le Professeur Maxime DOUGADOS (Paris)

Monsieur le Professeur Francis GUILLEMIN- Directeur de Thèse (Nancy)

Monsieur le Professeur Bernard MAZIERES- Rapporteur (Toulouse)

Monsieur le Professeur Philippe RAVAUD- Directeur de Thèse (Paris)

RAPPORTEURS : Monsieur le Professeur Bernard MAZIERES (Toulouse)

Monsieur le Professeur Gilles CHATELLIER (Paris)