

Différences méthodologiques dans les essais cliniques évaluant les traitements médicamenteux et les traitements non médicamenteux de la coxarthrose et de la gonarthrose

Isabelle Boutron, MD

Florence Tubach, MD

Bruno Giraudeau, PhD

Philippe Ravaud, MD, PhD

Il est largement admis que l'essai contrôlé randomisé (ECR) est la méthode la plus fiable pour déterminer l'efficacité d'un traitement spécifique.¹ Dans l'essai clinique, le plan expérimental, la conduite de l'étude et l'analyse ont pour objectifs de procurer des résultats valides, ce qui signifie que l'effet du traitement sera indiqué dans ses véritables direction et ampleur² et que les biais qui peuvent fausser les résultats seront minimisés.^{3,4}

L'arthrose de la hanche et celle du genou forment, à elles deux, une cause majeure d'invalidité.⁵ Elles peuvent être traitées soit par des méthodes non médicamenteuses, telles que la chirurgie, la rééducation, le lavage articulaire, l'acupuncture, les méthodes comportementales ou les cures thermales, soit par des traitements médicamenteux, tels que la prise d'un agent oral, une injection intra-articulaire ou un traitement topique.

Les essais contrôlés randomisés ont été développés essentiellement pour l'évaluation des traitements médicamenteux, probablement du fait de la pression exercée par les autorités réglementaires sur les laboratoires pharmaceutiques afin qu'ils effectuent des essais cliniques avant la mise sur le marché d'un nouveau médicament.⁶ Évaluer l'efficacité de traitements non médicamenteux pose des problèmes spécifiques.^{6,7} Les essais contrôlés randomisés sont mis en œuvre plus facilement pour évaluer les traitements médicamenteux parce que les investigateurs peuvent standardiser les posologies, mesurer l'observance thérapeutique, produire des placebos à l'identique et pratiquer les études en aveugle vis-à-vis des patients, des soignants et des personnes évaluant les résultats. Dans le cas des traitements non médicamenteux, il est souvent difficile, sur un plan technique ou éthique, de

Contexte Les essais contrôlés randomisés ont été développés essentiellement pour l'évaluation des traitements médicamenteux (c'est-à-dire, les médicaments pris par voie orale, les injections intra-articulaires, les traitements topiques et les traitements injectés par voie intramusculaire ou intraveineuse). L'évaluation de l'efficacité des traitements non médicamenteux (c'est-à-dire, les interventions chirurgicales, l'arthroscopie, le lavage articulaire, la rééducation, l'acupuncture et l'éducation des patients) pose, quant-à-elle, des problèmes spécifiques.

Objectifs Comparer la qualité des articles portant sur les traitements médicamenteux et sur les traitements non médicamenteux de la coxarthrose et de la gonarthrose et identifier les problèmes méthodologiques spécifiques à l'évaluation des traitements non médicamenteux.

Plan expérimental et cadre Nous avons recherché, dans Medline et dans le registre Cochrane des essais contrôlés, les articles portant sur les essais contrôlés randomisés évaluant des traitements médicamenteux et des traitements non médicamenteux chez des patients atteints d'arthrose de la hanche ou du genou, publiés entre le 1^{er} janvier 1992 et le 28 février 2002 dans 28 journaux de spécialité ou de médecine générale à facteur d'impact élevé.

Principaux critères de jugement La qualité des méthodes rapportées dans les articles sélectionnés a été évaluée, de manière indépendante, par 2 investigateurs, au moyen de l'échelle de Jadad, de la liste de Delphi et des recommandations indiquées dans le *Users' Guides to the Medical Literature*. Les investigateurs ont également utilisé une check-list d'items établie par les auteurs pour analyser les caractéristiques des études.

Résultats Cent dix articles au total ont été inclus dans l'analyse: 50 (45,5 %) évaluant des traitements non médicamenteux et 60 (54,5 %) des traitements médicamenteux. Les scores de qualité globale des rapports sur les traitements non médicamenteux ont été plus faibles que ceux des rapports sur les traitements médicamenteux, à la fois quand ils ont été mesurés par l'échelle de Jadad (score moyen [écart-type]: 1,4 [1,3] versus 3,0 [1,3]) et quand ils l'ont été par la liste de Delphi (score moyen [écart-type]: 5,2 [1,5] versus 7,5 [1,1]). Nous avons noté que les rapports ne faisaient pas toujours état d'une méthode correcte pour obtenir la séquence de randomisation et d'une analyse en intention de traiter, à la fois dans les articles portant sur les traitements non médicamenteux et dans ceux portant sur les traitements médicamenteux. Les interventions non médicamenteuses ont été moins souvent comparées à un placebo que les traitements médicamenteux (28,0 % des articles versus 71,7 %). Par comparaison avec les articles portant sur les traitements médicamenteux, ceux sur les interventions non médicamenteuses ont indiqué moins souvent que les patients (26,0 % versus 96,7 %), les soignants (6,0 % versus 81,7 %) et les personnes ayant évalué les résultats (68,0 % versus 98,3 %) étaient dans l'ignorance du traitement attribué par randomisation. Le niveau de compétences des soignants pouvait influencer l'effet du traitement dans 84,0 % des cas dans les articles traitant des interventions non médicamenteuses contre 23,3 % des cas dans ceux portant sur les traitements médicamenteux.

Conclusions Dans cette analyse de rapports sur les traitements de la coxarthrose et de la gonarthrose, les articles portant sur les interventions non médicamenteuses ont eu des scores plus faibles, en terme de qualité, que les articles portant sur les traitements médicamenteux. L'évaluation des interventions non médicamenteuses doit prendre en considération des problèmes supplémentaires de méthodologie.

JAMA.2003; 290: 1062-1070.

www.jama.com

Affiliations des auteurs: Département d'Epidémiologie, Biostatistique et Recherche Clinique, Groupe Hospitalier Bichat – Claude Bernard (Assistance Publique des Hôpitaux de Paris), Faculté Xavier Bichat, Université Paris VII (Drs Boutron, Tubach, and Ravaud) and INSERM EMI 0357 (Drs Boutron, Tubach, and Ravaud), Paris, France;

and INSERM CIC 202, Tours, France (Dr Giraudeau).

Correspondance: Isabelle Boutron, MD, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Groupe Hospitalier Bichat – Claude Bernard, 46 rue Henri Huchard, 75877 Paris, Cedex 18 France (e-mail: isabelle.boutron@bch.ap-hop-paris.fr).

pratiquer une intervention factice et il est souvent impossible de laisser les patients et les soignants dans l'ignorance du traitement attribué, alors que l'effet placebo de ce type de traitement est probablement important. Examinons un exemple. Des rapports d'ECR évaluant le lavage articulaire^{8,9} ont décrit l'efficacité de ce traitement dans la gonarthrose. Cependant, ces études ont été effectuées sans qu'une intervention factice n'ait été pratiquée dans le groupe témoin et alors que les patients n'ignoraient pas le traitement qu'ils allaient recevoir. Ces résultats sont en contradiction avec ceux d'un autre ECR évaluant le lavage articulaire, cet essai ayant été effectué en aveugle vis-à-vis des patients et des personnes chargées d'évaluer les résultats et une intervention factice ayant été pratiquée dans le groupe témoin.¹⁰ Ces résultats contradictoires peuvent être reliés soit à la pratique d'une intervention dans le groupe témoin soit à la variabilité des résultats. De plus, contrairement à ce qui se passe pour les traitements médicamenteux, dans le cas d'une intervention non médicamenteuse le soignant fait partie intégrante du traitement ; le succès du traitement dépend des compétences de celui-ci, de son expérience et de son enthousiasme. Il est, par exemple, bien connu que les résultats d'arthroplastie du genou ou de la hanche dépendent de l'expérience du chirurgien et du volume d'interventions pratiquées dans l'hôpital.^{11,12} Enfin, les traitements non médicamenteux sont généralement complexes et difficiles à standardiser, et des modifications techniques peuvent survenir au fur et à mesure que la technique évolue. Ces problèmes méthodologiques ne sont généralement pas pris en compte dans l'appréciation de la qualité des articles évaluant un traitement non médicamenteux.

Les objectifs de cette étude ont été de comparer la qualité de rapports effectués sur des traitements médicamenteux et des traitements non médicamenteux de la coxarthrose et de la gonarthrose et d'identifier les problèmes méthodologiques spécifiques à l'évaluation des traitements non médicamenteux.

MÉTHODES

Stratégie de recherche

Nous avons effectué des recherches, dans Medline et dans le registre Cochrane des essais contrôlés, en utilisant les termes suivants : *osteoarthritis* (arthrose) OU *osteoarthritic* (arthrosique) et *hip* (hanche) OU *knee* (genou), en nous limitant à : *clinical trials* (essais cliniques). Nous avons identifié et sélectionné tous les compte-rendus d'ECR évaluant les traitements non médicamenteux et les traitements médicamenteux des patients atteints d'arthrose de hanche ou du genou, publiés entre le 1^{er} janvier 1992 et le 28 février 2002 dans les journaux suivants (en nous basant sur leur facteur d'impact indiqué pour l'année 2001) :

(1) les 10 journaux de médecine interne et de médecine générale ayant le facteur d'impact le plus élevé (*New England Journal of Medicine*, *JAMA*, *The Lancet*, *Annals of Internal Medicine*, *Annual Review of Medicine*, *Archives of Internal Medicine*, *BMJ*, *American Journal of Medicine*, *Medicine*, et *Proceedings of the Association of American Physicians*) ; (2) les 6 journaux de rhumatologie ayant le facteur d'impact le plus élevé (*Arthritis and Rheumatism*, *Seminars in Arthritis and Rheumatism*, *Annals of the Rheumatic Diseases*, *Rheumatology [Oxford, Royaume-Uni]*, *Journal of Rheumatology*, et *Rheumatic Diseases Clinics of North America*) ; (3) les 6 journaux d'orthopédie ayant le facteur d'impact le plus élevé (*Osteoarthritis and Cartilage/OARS*, *Osteoarthritis Research Society* ; *Journal of Orthopaedic Research : Official Publication of the Orthopaedic Research Society* ; *Journal of Bone and Joint Surgery, American Volume* ; *Spine* ; *Gait and Posture* ; et *Journal of Bone and Joint Surgery, British Volume*) ; et (4) les 6 journaux de rééducation ayant le facteur d'impact le plus élevé (*Archives of Physical Medicine and Rehabilitation*, *Supportive Care in Cancer : Official Journal of the Multinational Association of Supportive Care in Cancer*, *Journal of Electromyography and Kinesiology : Official Journal of the International Society of Electrophysiological Kinesiology*, *Physical Therapy*, *Journal of Rehabilitation Research and Development*, et *Scandinavian Journal of Rehabilitation Medicine*).

Nous avons choisi ces journaux parce qu'un facteur d'impact élevé est prédictif d'une qualité méthodologique élevée des articles du journal en question¹³ et que notre but était non pas d'être exhaustifs mais plutôt de faire prendre conscience des problèmes méthodologiques existant lors de l'évaluation de traitements non médicamenteux.

Les articles recueillis ont été évalués par l'une d'entre nous (I. Boutron), qui a examiné les titres et les résumés afin d'identifier les études pertinentes. Les articles n'ont été inclus que si l'étude était identifiée comme étant un ECR, si l'article était publié dans son intégralité et si l'objet de l'étude était l'évaluation d'un traitement médicamenteux ou d'un traitement non médicamenteux de la coxarthrose ou de la gonarthrose. Les études non contrôlées, portant sur des séries de cas, et les articles publiés uniquement sous forme de résumé ou d'éditorial ou dans les rubriques actualités et correspondance ont été exclues. Les publications en double ont été recherchées (c'est-à-dire, le même essai publié 2 fois avec des résultats distincts provenant de différentes durées de suivi) ; dans ce cas, seul l'article le plus récent était sélectionné en vue d'inclusion.

Evaluation de la qualité des études

La qualité méthodologique des articles sélectionnés a été évaluée, de manière indépendante, par 2 investigatrices (I. Boutron

et F. Tubach), au moyen de l'échelle de Jadad¹⁴ et de la liste de Delphi,¹⁵ ces méthodes ayant été utilisées, car étant validées et largement répandues. L'*Users' Guides to the Medical Literature*¹⁶ a également été utilisé. Cependant, à notre connaissance, il n'existe aucun outil d'évaluation de la qualité qui soit spécifique des traitements non médicamenteux. Les articles ont donc été évalués également en utilisant une check-list d'items élaborés par les auteurs, afin de cibler les problèmes méthodologiques qui existent lorsque l'on évalue un traitement non médicamenteux (cette liste est disponible en ligne, à l'adresse suivante : <http://www.bichat.inserm.fr/fichierpdf/emi0357/checklistofitems1.pdf>). Cette check-list collectait les informations suivantes : année de publication, sources de financement (fonds publics ou privés), nombre de centres impliqués, type de traitement évalué (médicament pris par voie orale, traitement topique, injection intra-articulaire, intervention chirurgicale, arthroscopie, lavage articulaire, acupuncture, rééducation ou méthode comportementale), et si les éléments du graphique CONSORT (*Consolidated Standards of Reporting Trials*, *regroupement des normes relatives aux rapports d'études*)¹⁷ étaient indiqués dans un ordigramme ou dans un texte. Les informations concernant le plan expérimental de l'étude et le mode de randomisation ont également été collectées et il a été vérifié si la méthode de dissimulation de la répartition des traitements était correcte. Le procédé utilisé pour obtenir la séquence de randomisation a été considéré comme adéquat si une prévention des biais de sélection avait été faite en utilisant, par exemple, une table de nombres au hasard, des nombres au hasard générés par ordinateur ou un tirage au sort à pile ou face ou par cartes. La répartition des traitements a été considérée comme correctement dissimulée si les patients et les personnes qui avaient recruté les patients ne pouvaient découvrir quel traitement était attribué du fait de l'utilisation, par exemple, d'une centralisation de la randomisation, d'un contrôle de la pharmacie, d'enveloppes opaques et cachetées ou de flacons ou récipients numérotés ou codés.

Les investigatrices ont examiné si le traitement était individualisé, ce qui supposait une modification de la posologie ou du mode de traitement en fonction de la tolérance du patient ou de la comorbidité et en fonction de l'efficacité clinique. Elles ont noté si, selon elles, l'intervention était décrite avec suffisamment de détails pour être reproductible. Dans les situations pouvant impliquer une influence du soignant sur le succès du traitement, les investigatrices ont analysé si la courbe d'apprentissage et si l'expérience du soignant étaient prises en compte et si les soignants étaient qualifiés.

L'observance thérapeutique a été évaluée dans le cas des interventions répétées (c'est-à-dire, les traitements nécessitant des actions

répétées, telles que des prises médicamenteuses ou des séances de kinésithérapie, contrairement à la chirurgie). Observances quantitative et qualitative ont été différenciées. L'observance quantitative évaluait, par exemple, si un patient avait effectué toutes les séances de kinésithérapie ou pris tous les médicaments oraux prescrits ; l'observance qualitative évaluait si le patient avait effectué l'intervention de manière correcte (par exemple, si les exercices effectués à domicile par le patient étaient en accord avec le programme prescrit). La méthode utilisée pour mesurer l'observance (par exemple, la vérification de la posologie du médicament, le comptage des comprimés, l'interrogatoire du patient ou la tenue d'un journal) était consignée.

Les investigatrices ont examiné quelle intervention avait été effectuée dans le groupe témoin et si l'effet placebo potentiel de chacune des interventions était, selon elles, similaire. Par exemple, elles ont considéré comme similaires l'effet placebo de l'utilisation d'un anti-inflammatoire non stéroïdien et celui d'un placebo présenté de façon identique. L'effet placebo d'un lavage articulaire effectué dans des conditions aseptiques, en salle d'opérations, sous anesthésie locale et avec injection d'1 litre de solution saline, et celui d'une seule injection intra-articulaire sont, cependant, probablement différents. Les informations concernant les similitudes de traitements concomitants dans chaque groupe (thérapeutique supplémentaire imprévue dans chacun des groupes de comparaison) ont également été prises en considération, ainsi que celles qui concernaient la survenue ou la probabilité d'une contamination entre les groupes de traitement (intervention étudiée effectuée dans le groupe témoin).

Les investigatrices ont également étudié si les informations suivantes étaient indiquées : l'essai avait-il été effectué en aveugle pour les patients, les soignants et les personnes évaluant les résultats, et des tests permettant de vérifier la réalité de l'ignorance du traitement attribué avaient-ils été pratiqués ? Effectuer une étude en aveugle signifie que l'on garde les participants, les soignants et les personnes évaluant les résultats dans l'ignorance du traitement attribué, afin qu'aucun d'entre eux ne soit influencé par la connaissance de ce traitement.¹⁸

Les investigatrices ont examiné si le calcul du nombre de sujets nécessaires à l'essai, ou si une évaluation a posteriori de la puissance de l'essai, était indiqué et si une analyse en intention de traiter¹⁹ était rapportée et/ou effectuée. Enfin, elles se sont efforcées de suivre l'approche du *Users' Guides to the Medical Literature*¹⁶ pour juger de manière subjective la validité de l'étude – selon ce manuel, l'évaluation finale de la validité ne se traduit jamais par une décision de "valide" ou "non valide", mais il existe plutôt un continuum entre des études de haute qualité ayant évité les biais et des études de faible qualité qui probablement fournissent une estimation biaisée des résultats.

Une évaluation subjective de la qualité des études a donc été effectuée en utilisant une échelle de jugement numérique allant de 1 à 10, après réponse à la question suivante : "jusqu'à quel point les erreurs systématiques ou les biais ont-ils été évités dans ce rapport ?" Etant donné la subjectivité de cette évaluation globale, nous ne pouvons exclure un biais à l'encontre des études concernant les traitements non médicamenteux.

Avant d'entreprendre ce travail, les 2 investigatrices ont effectué une évaluation sur une série distincte de 10 articles. Puis, elles ont discuté en réunion de l'interprétation des différentes échelles, afin de résoudre toute différence dans l'établissement des scores. Durant l'étude, chacune d'entre elles a examiné, d'une manière indépendante, les articles sélectionnés dans une séquence aléatoire, générée par ordinateur, différente. Elles ont évalué le titre, ainsi que les rubriques "méthodes" et "résultats". Le nom du journal et les auteurs ne leur ont pas été dissimulés, puisque les données concernant les effets de cette dissimulation, dans le cadre des évaluations de la qualité des essais, ne sont pas cohérentes entre elles.^{20,21} Les divergences d'évaluation des articles sélectionnés entre les 2 investigatrices ont été résolues par consensus. Pour chaque item divergent, l'article était lu de nouveau par chacune d'entre elles pour parvenir à un accord. Les données que nous rapportons ici sont issues de ce consensus. La qualité des articles sélectionnés a été déterminée en utilisant la moyenne de l'appréciation globale de la qualité des articles donnée par chacune des 2 personnes ayant fait l'évaluation. Cette appréciation provenait de scores établis sur une échelle de jugement numérique (score allant de 1 à 10), sur l'échelle de Jadad (score allant de 0 à 5) et sur la liste de Delphi (score de qualité globale se composant du nombre d'items satisfaits et allant de 0 à 9).²² Pour toutes les échelles, un score élevé indique une qualité élevée.

Analyse statistique

Les statistiques descriptives (moyennes, écarts-types, valeurs minimales et valeurs maximales) ont été utilisées pour les variables continues. Les variables qualitatives ont été décrites au moyen de fréquences et de pourcentages. Le degré d'accord entre les 2 investigatrices a été établi en utilisant le coefficient κ pour les variables qualitatives. La fiabilité inter-investigatrices a été évaluée en utilisant le coefficient de corrélation intra-classe (CCI) pour les variables continues. Toutes les analyses ont été effectuées au moyen du logiciel SAS version 8.2 (SAS Institute Inc., Cary, Caroline du Nord).

RÉSULTATS

Articles sélectionnés

Parmi les 198 articles identifiés, 119 ont été sélectionnés en vue de leur évaluation (voir

Figure 1). Parmi les 79 articles exclus, 5 étaient des résumés uniquement et 3 des publications en double ; dans 12 cas, l'étude n'était pas un ECR, dans 45 cas, elle n'évaluait pas une intervention thérapeutique, dans 13 cas, elle n'évaluait pas un traitement de la coxarthrose ou de la gonarthrose et, dans 1 cas, il s'agissait d'un essai de phase II. Neuf articles ont été exclus, dans un deuxième temps, après l'obtention du texte intégral, soit parce qu'il n'y avait pas d'ECR ($n = 3$), soit parce qu'il s'agissait d'analyses de sous-groupes ($n = 1$), soit parce qu'un suivi plus long du même ECT était décrit dans un autre article ($n = 5$). Cent dix articles ont donc été inclus dans notre analyse. Parmi eux, 50 (45,5 %) évaluaient des traitements non médicamenteux : intervention chirurgicale ($n = 15$), lavage sous arthroscopie ($n = 3$), lavage articulaire ($n = 3$), rééducation ($n = 23$), éducation du patient ($n = 2$), cure thermique ($n = 1$) et acupuncture ($n = 1$). Les interventions de rééducation comprenaient : kinésithérapie ($n = 14$), appareillage orthopédique ($n = 5$), et stimulation électrique nerveuse percutanée ou thérapie par le laser ($n = 4$). Soixante articles (54,5 %) évaluaient des traitements médicamenteux : médicaments administrés par voie orale ($n = 41$), locale ($n = 3$), intraveineuse ($n = 1$) ou intramusculaire ($n = 1$) ; ou injections intra-articulaires ($n = 14$).

La fiabilité inter-investigatrices a été bonne concernant la méthode utilisée pour obtenir la séquence de randomisation (accord : 92,7 % ; $\kappa = 0,86$; intervalle de confiance [IC] à 95 % : 0,77 – 0,95), la dissimulation de la répartition des traitements (accord : 89,1 % ; $\kappa = 0,79$; IC à 95 % : 0,67 – 0,90), l'ignorance du patient (accord : 96,4 % ; $\kappa = 0,92$; IC à 95 % : 0,85 – 1,00), celle du soignant (accord : 96,4 % ; $\kappa = 0,93$; IC à 95 % : 0,86 – 1,00) et celle de la personne évaluant les résultats (accord : 97,3 % ; $\kappa = 0,90$; IC à 95 % : 0,78 – 1,00) concernant le traitement attribué. Déterminer si l'analyse avait été faite en intention de traiter a abouti à une valeur de κ plus faible (accord : 70,9 % ; $\kappa = 0,59$; IC à 95 % : 0,48 – 0,71). La fiabilité inter-investigatrices, évaluée en utilisant le CCI, a été bonne pour l'échelle de Jadad (CCI : 0,84 ; IC à 95 % : 0,78 – 0,89), la liste de Delphi (CCI : 0,88 ; IC à 95 % : 0,83 – 0,92) et l'échelle de jugement numérique (CCI : 0,62 ; IC à 95 % : 0,49 – 0,72).

Seulement 10 articles (9,1 %) ont été publiés dans un journal de médecine générale (*JAMA*, *The Lancet*, *BMJ*, *Annals of Internal Medicine* et *Archives of Internal Medicine*). Les autres articles ont été publiés principalement dans les journaux suivants : *Journal of Rheumatology* (20,9 %), *Osteoarthritis and Cartilage* (16,4 %), *Arthritis and Rheumatism* (11,8 %), *Journal of Bone and Joint Surgery*, *British Volume* (11,8 %), *Rheumatology* (10,0 %) et *Annals of the Rheumatic Diseases* (9,1 %).

Les sources de financement étaient privées, en totalité ou en partie, dans 57 articles (51,8 %) et publiques dans 26 articles (23,6 %) ;

elles n'étaient pas mentionnées dans 27 articles (24,6 %). Concernant les traitements médicamenteux, les fonds étaient principalement indiqués comme étant d'origine privée (75 %) ou bien leur provenance n'était pas mentionnée (23,3 %) ; pour les traitements non médicamenteux, les fonds étaient indiqués comme étant d'origine publique dans 25 articles (50,0 %) et privée dans 12 articles (24,0 %), leur provenance n'étant pas mentionnée dans 13 articles (26,0 %). La moitié des articles concernaient des essais multicentriques, ces derniers étant plus fréquents en cas de traitement médicamenteux qu'en cas de traitement non médicamenteux (68,3 % versus 30,0 %).

Les différents éléments du graphique CONSORT (flux de participants à chaque étape de l'essai) ont été indiqués soit dans le texte soit dans un ordigramme dans 48 des 75 articles publiés depuis le communiqué CONSORT en 1996.²³ Ces données ont été indiquées dans 71,4 % (30/42) des articles sur les traitements médicamenteux et dans 54,5 % (18/33) des articles portant sur les interventions non médicamenteuses (seulement dans 9,1 % [1/11] pour les articles traitant de chirurgie).

Evaluation de la qualité

Quel que soit l'outil d'évaluation utilisé, les scores de qualité ont été meilleurs pour les articles portant sur les traitements médicamenteux que pour ceux portant sur les traitements non médicamenteux (voir Figure 2). Sur l'échelle de Jadad, les articles sur les traitements médicamenteux ont eu des scores moyens (écart-type) de 3,0 (1,3) contre 1,4 (1,3) pour les articles sur les traitements non médicamenteux. La connaissance du traitement attribué dans le cas des interventions non médicamenteuses explique en grande partie cette différence. Sur la liste de Delphi, les articles sur les traitements médicamenteux ont eu un score moyen (écart-type) de 7,5 (1,1), contre 5,2 (1,5) pour les articles sur les traitements non médicamenteux. Sur l'échelle de jugement numérique, les scores moyens ont été respectivement de 7,0 (1,7) et de 4,9 (2,0). En outre, comme on peut le voir sur la Figure 2, les scores de qualité les moins bons ont été attribués aux articles traitant de chirurgie/arthroscopie/lavage articulaire ; les scores ont indiqué une faible qualité, similaire, pour les rapports sur la rééducation et ceux sur les injections intra-articulaires ; les scores les plus élevés ont été attribués aux articles sur les autres traitements médicamenteux. Des informations complémentaires sont disponibles en ligne, à l'adresse suivante : <http://www.bichat.inserm.fr/fichierpdf/em0357/dataonarticlesassessed.pdf>.

Caractéristiques des essais

Plans expérimentaux des études. Les études ont été réalisées en parallèle dans 104 cas ; elles étaient de type croisé dans 6 cas. La sélection

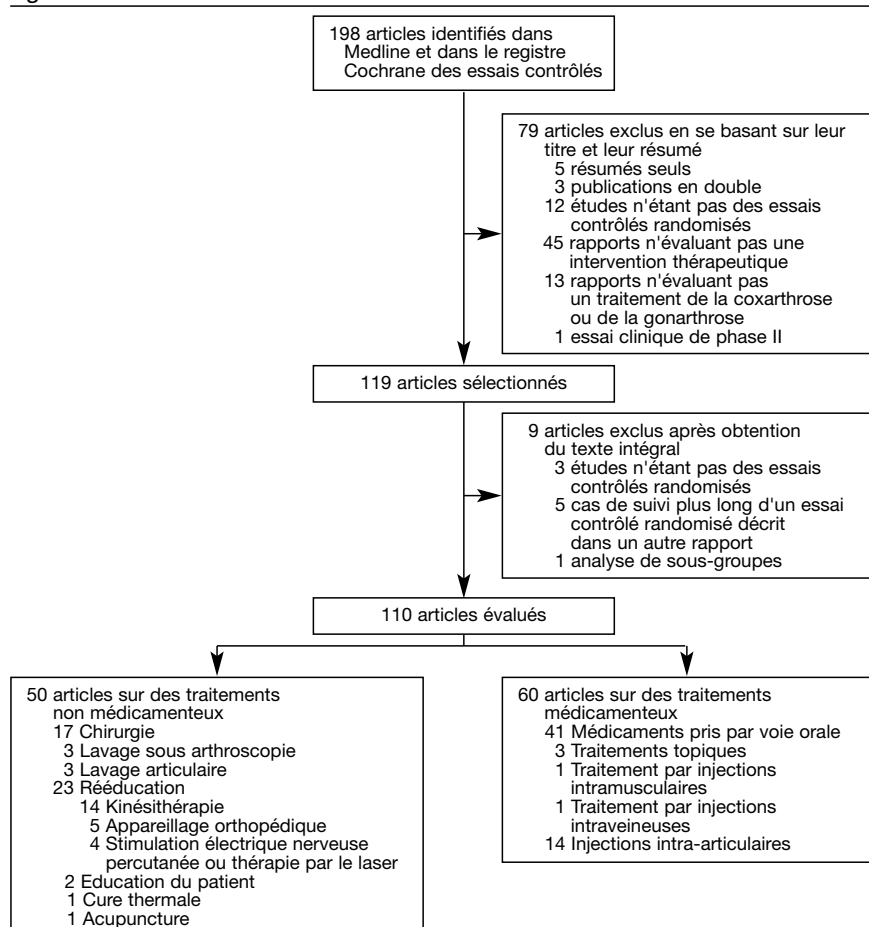
des articles impliquait dans tous les cas une randomisation des patients. Cependant, la méthode utilisée pour obtenir la séquence de randomisation n'était correcte que dans 49,1 % des articles et cette séquence était dissimulée aux investigateurs ayant recruté les patients dans 20,9 % des cas seulement (voir Tableau 1). Aucune différence n'a été trouvée entre les articles sur les traitements médicamenteux et ceux sur les traitements non médicamenteux concernant l'indication d'une méthode correcte pour obtenir la séquence de randomisation. Le nombre de rapports indiquant une dissimulation correcte de la répartition des traitements a été similaire dans les 2 cas. Cependant la dissimulation était plus souvent incorrecte dans les articles sur les traitements non médicamenteux que dans ceux sur les traitements médicamenteux, tandis qu'il y avait plus souvent une absence de compte-rendu de la dissimulation dans le cas des articles sur les traitements médicamenteux (voir Tableau 1).

Description des interventions. Reproductibilité et individualisation. L'intervention a été plus souvent décrite avec suffisamment de

détails pour être reproductible dans les articles sur les traitements médicamenteux que dans ceux sur les traitements non médicamenteux (voir Tableau 1). Parmi ces derniers, ce sont les traitements chirurgicaux qui ont été considérés comme le moins souvent reproductibles (voir Tableau 2). L'intervention n'était presque jamais individualisée dans les articles sur les traitements médicamenteux, tandis qu'elle était décrite comme telle dans un tiers des cas dans les articles sur les traitements non médicamenteux (voir Tableau 1). Enfin, dans les articles sur les traitements non médicamenteux, la qualité de la technique de traitement n'a jamais été évaluée.

Soignants. Le niveau de compétences et l'expérience des soignants pouvaient influencer l'effet du traitement, en particulier quand ce traitement était une intervention chirurgicale quelle qu'elle soit, dans la plupart des articles sur les traitements non médicamenteux, mais ces facteurs n'avaient d'influence que dans le cas des injections intra-articulaires dans les articles sur les traitements médicamenteux (voir Tableaux 1 et 2). Contrastant avec ces faits,

Figure 1. Processus de sélection des articles de l'étude



DIFFÉRENCES DANS LES ESSAIS THÉRAPEUTIQUES DE L'ARTHROSE

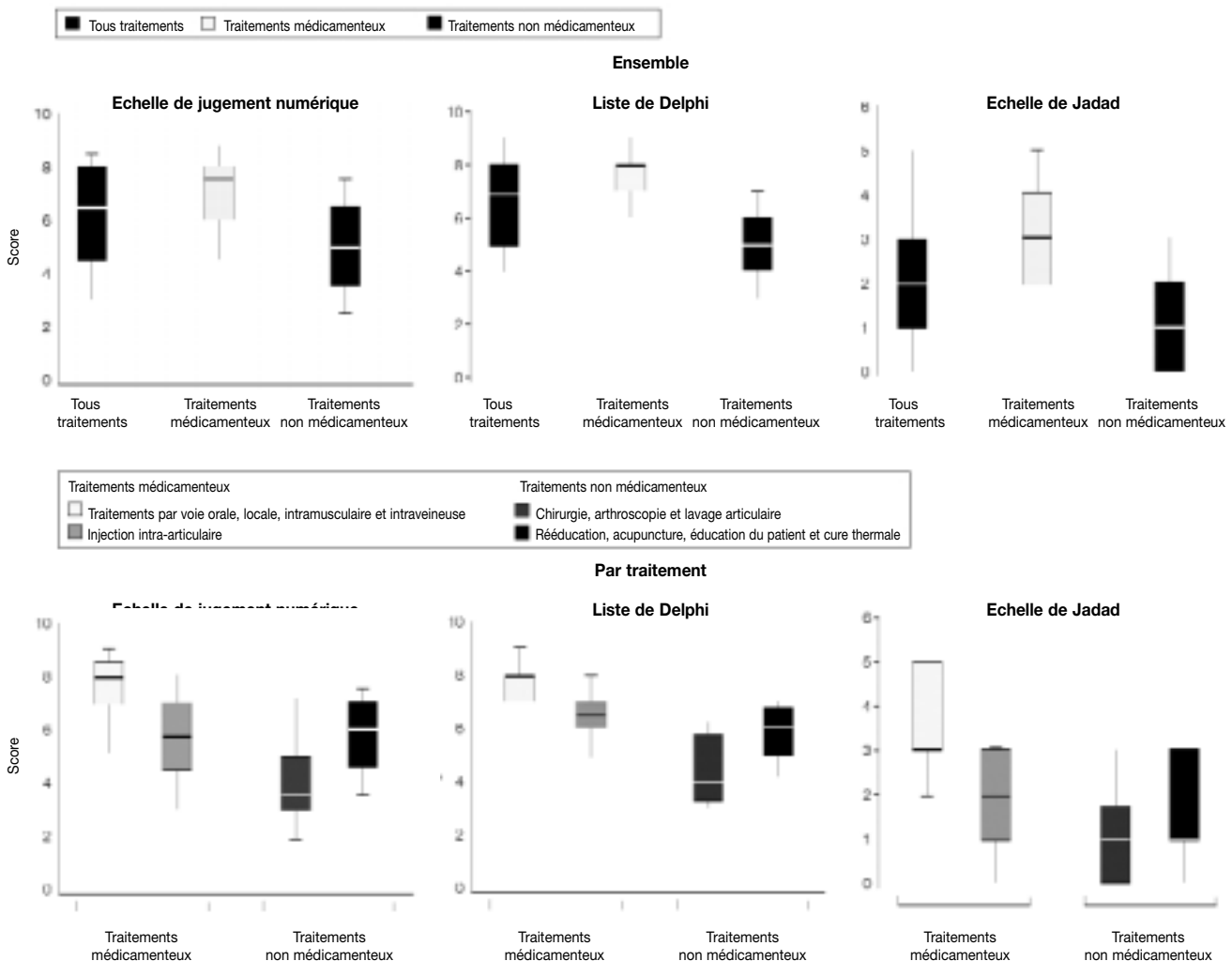
l'expérience du soignant n'a été indiquée que dans 6 articles seulement et le volume d'interventions pratiquées dans l'hôpital dans 1 seul article ; la courbe d'apprentissage des soignants, quant à elle, n'a jamais été prise en compte. Enfin, la formation du soignant avant le début de l'essai n'a été mentionnée que dans 2 articles qui évaluaient, pour l'un, les injections intra-articulaires et, pour l'autre, la rééducation.

Observance thérapeutique. Nous n'avons étudié l'observance thérapeutique que pour les interventions réitérées (c'est-à-dire, les traitements nécessitant des actions répétées telles que des prises médicamenteuses ou des séances de kinésithérapie), qui faisaient l'objet de 57 (95,0 %) articles sur les traitements médicamenteux et de 26 (52,0 %) articles sur les traitements non médicamenteux. Parmi ces

articles, le nombre de comptes-rendus concernant l'observance thérapeutique a été similaires dans ceux sur les traitements médicamenteux (32/57, soit 56,1 %) et dans ceux sur les traitements non médicamenteux (16/26, soit 61,5 %). La qualité de l'observance n'a, cependant, jamais été évaluée dans ces articles. L'observance a été déterminée par au moins un critère objectif (comptage des comprimés ou vérification de la posologie du médicament) dans tous les articles sur les traitements médicamenteux et 7 fois sur 16 (43,7 %) dans les articles sur les traitements non médicamenteux (nombre de séances suivies pour la kinésithérapie ou dispositif enregistrant les heures d'utilisation pour la stimulation électrique nerveuse percutanée).

Intervention dans le groupe témoin. L'intervention témoin rapportée dans les articles sur les traitements médicamenteux était plus souvent un placebo, dans au moins 1 groupe. Dans les articles sur les traitements non médicamenteux, le groupe subissant le traitement à évaluer était plus souvent comparé à un groupe témoin ayant un traitement actif ou le traitement habituel ou à un groupe témoin de personnes placées sur liste d'attente d'une intervention active (voir Tableau 1). Les interventions chirurgicales ont toujours été comparées à une intervention témoin active et jamais à un placebo (voir Tableau 2). L'effet placebo potentiel des différents traitements comparés a été plus souvent considéré comme similaire

Figure 2. Evaluation de la qualité des rapports d'essais effectués sur des traitements médicamenteux et sur des traitements non médicamenteux.



Les scores sont basés sur une échelle de jugement numérique, s'étendant de 1 à 10; sur la liste de Delphi, ²² s'étendant de 0 à 9; et sur l'échelle de Jadad, ¹⁴ s'étendant de 0 à 5. Des scores élevés correspondent à une haute qualité. Les rectangles représentent les scores des données étudiées, de la médiane (trait horizontal) jusqu'aux 25^e (sommets du rectangle) et 75^e (base du rectangle) percentiles. Dans quelques cas, l'observation médiane a coïncidé avec le 25^e ou le 75^e percentile. Les barres d'erreurs indiquent les 10^e et 90^e percentiles.

dans les articles portant sur les traitements médicamenteux (voir Tableau 1).

Traitements concomitants. La description des traitements concomitants (c'est-à-dire les traitements supplémentaires, distincts des interventions, apportés dans chaque groupe de comparaison) a été rapportée plus souvent dans les articles sur les traitements médicamenteux (58,3 %) que dans ceux sur les traitements non médicamenteux (24,0 %). L'existence d'une contamination entre groupes n'a été indiquée que dans 3 articles.

Ignorance du traitement attribué. Dans les articles sur les traitements médicamenteux, il a quasiment toujours été mentionné que les patients étaient dans l'ignorance du traitement attribué, mais cela n'a été le cas que dans un quart des cas dans les articles sur les traitements non médicamenteux (voir Tableau 1). S'il était indiqué que les soignants n'avaient pas connaissance du traitement attribué dans 81,7 % des articles sur les traitements médicamenteux, cela a été rarement le cas dans les articles sur les traitements non médicamenteux (voir Tableau 1). Dans les cas où il était indiqué que les patients et les soignants étaient dans l'ignorance du traitement attribué, la réalité de cette ignorance n'a jamais été testée chez les soignants et elle n'a été testée que dans 1 étude sur un traitement médicamenteux et dans 1 étude sur un traitement non médicamenteux. Quand les patients connaissaient leur traitement, dans seulement 2 articles sur les traitements non médicamenteux, il leur a été demandé de ne pas communiquer cette information à la personne chargée d'évaluer les résultats. Enfin, la personne chargée de l'évaluation était moins souvent dans l'ignorance du traitement attribué dans les articles portant sur les traitements non médicamenteux (voir Tableau 1).

Evaluation des résultats. Dans seulement 3 articles sur les traitements non médicamenteux et 4 articles sur les traitements médicamenteux, les personnes chargées de l'évaluation ont été formées et il n'a jamais été indiqué qu'une évaluation des résultats ait été faite par un comité chargé de revoir les critères de jugement (un comité indépendant qui décide en dernier ressort si un participant remplit les critères pour un des résultats de l'étude). L'apparition d'effets indésirables a été rapportée plus souvent dans les articles sur les traitements médicamenteux que dans ceux sur les traitements non médicamenteux (83,3 % versus 46,0 %). Les effets indésirables ont été évalués par un comité indépendant dans seulement 2 articles.

Analyses en intention de traiter et justification de la taille de l'échantillon. Les analyses statistiques ont été effectuées selon le principe de l'intention de traiter²⁴ (c'est-à-dire que tous les participants randomisés sont inclus dans l'analyse et gardés dans leur groupe d'origine) dans seulement 30 % de

l'ensemble des articles (voir Tableau 1). Un calcul du nombre de sujets nécessaires à l'étude, ou une estimation a posteriori de la puissance de l'étude, a été rapporté plus souvent dans les articles sur les traitements médicamenteux que dans ceux sur les traitements non médicamenteux, et a été particulièrement rarement indiqué dans les articles portant sur la chirurgie.

Influence du journal

Nous avons analysé "l'effet journal" et découvert que les articles publiés dans les journaux de médecine interne et médecine générale (n = 10), ainsi que ceux publiés dans les journaux de rhumatologie (n = 57), avaient un score de qualité plus élevé que ceux publiés dans les journaux d'orthopédie (n = 36) et de rééducation (n = 7), en nous fondant sur les scores obtenus sur l'échelle de Jadad (score moyen [écart-type] : 2,8 [1,3] versus 1,5 [1,4]), sur la liste de Delphi (7,0 [1,2] versus 5,5 [1,9]) et sur l'échelle de jugement numérique (6,7 [1,8] versus 5,1 [2,1]).

COMMENTAIRES

Cette étude a évalué la qualité méthodologique de tous les ECR portant sur les traitements de la coxarthrose et de la gonarthrose, publiés sur une période de 10 ans dans les journaux de médecine générale et de spécialité ayant un facteur d'impact élevé.

Plusieurs travaux ont évalué la qualité méthodologique d'une large gamme de rapports d'essais randomisés, dans différents domaines de la santé. Quel que soit le domaine évalué, la qualité globale des ECR n'est pas bonne. Il est fréquent de constater des problèmes méthodologiques au niveau de la randomisation et de l'analyse en intention de traiter.^{1,25,26} De plus, les résultats de ces travaux montrent que de tels défauts méthodologiques peuvent influencer l'importance de l'effet du traitement. Il a été montré que si une étude a une méthode d'obtention de la séquence de randomisation incorrecte, si cette séquence n'est pas dissimulée et si l'étude n'est pas effectuée en double-aveugle, l'effet du traitement est fortement augmenté.^{3,24,27} Cependant, à notre connaissance,

Tableau 1. Caractéristiques des essais contrôlés randomisés évaluant les traitements médicamenteux et les traitements non médicamenteux de la coxarthrose et de la gonarthrose*.

Caractéristiques	Essais évaluant les traitements non médicamenteux (n = 50)	Essais évaluant les traitements médicamenteux (n = 60)	Ensemble des essais (n = 110)
Randomisation			
Méthode utilisée pour obtenir la séquence de randomisation			
Correcte	27 (54,0)	27 (45,0)	54 (49,1)
Incorrecte	2 (4,0)	0	2 (1,8)
Non indiquée	21 (42,0)	33 (55,0)	54 (49,1)
Dissimulation de la répartition des traitements			
Correcte	9 (18,0)	14 (23,3)	23 (20,9)
Incorrecte	15 (30,0)	1 (1,7)	16 (14,5)
Non indiquée	26 (52,0)	45 (75,0)	71 (64,6)
Reproductibilité de l'intervention	41 (82,0)	59 (98,3)	100 (90,9)
Individualisation	15 (30,0)	2 (3,3)	17 (15,4)
Influence des compétences du soignant	42 (84,0)	14 (23,3)	56 (50,9)
Intervention dans le groupe témoin			
Placebo	14 (28,0)	43 (71,7)	57 (51,8)
Traitement actif chez les témoins	28 (56,0)	16 (26,7)	44 (40,0)
Traitement habituel ou groupe témoin sur liste d'attente	8 (16,0)	1 (1,7)	9 (8,2)
Effet placebo potentiellement similaire	35 (70,0)	59 (98,3)	94 (85,4)
Ignorance du traitement attribué			
Patients	13 (26,0)	58 (96,7)	71 (64,5)
Soignants	3 (6,0)	49 (81,7)	52 (47,3)
Personnes évaluant les résultats	34 (68,0)	59 (98,3)	93 (84,5)
Analyse			
AIT effectuée en totalité†	12 (24,0)	21 (35,0)	33 (30,0)
AIT modifiée‡	10 (20,0)	19 (31,7)	29 (26,4)
Analyse per-protocole effectuée	24 (48,0)	17 (28,3)	41 (37,3)
Peu clair	4 (8,0)	3 (5,0)	7 (6,3)
Indication du calcul du nombre de sujets à inclure ou de la puissance de l'étude	15 (30,0)	42 (70,0)	57 (51,8)

Abréviations : AIT = analyse en intention de traiter.

* Toutes les données sont exprimées en nombre (pourcentage).

† Tous les participants randomisés ont été inclus dans l'analyse selon le groupe qui leur a été attribué à l'origine par le tirage au sort.

‡ Les analyses ont exclu les participants n'ayant jamais pris leur traitement ou n'ayant jamais été évalués alors qu'ils prenaient leur traitement.

DIFFÉRENCES DANS LES ESSAIS THÉRAPEUTIQUES DE L'ARTHROSE

aucune étude n'a comparé la qualité méthodologique des articles sur les traitements médicamenteux et celle des articles sur les traitements non médicamenteux. Nous nous sommes concentrés sur la qualité méthodologique des articles portant sur les ECR évaluant les traitements de la coxarthrose et la gonarthrose car ceux-ci couvrent une vaste gamme de traitements non médicamenteux (par exemple, chirurgie, arthroplastie, lavage articulaire, gymnastique médicale, kinésithérapie, port d'orthèse, cure thermique, acupuncture et éducation du patient) et médicamenteux (par exemple, médicaments pris par voie orale, injection intra-articulaire).

Notre analyse a montré que la qualité méthodologique des rapports d'essais évaluant des traitements non médicamenteux était inférieure à celle des rapports d'essais évaluant des traitements médicamenteux. Des défauts méthodologiques, tels que l'absence d'indications d'un procédé correct d'obtention de la

séquence de randomisation, d'une dissimulation correcte de la répartition des traitements et d'une analyse faite en intention de traiter, ont été fréquents, à la fois dans les articles portant sur les traitements non médicamenteux et dans ceux portant sur les traitements médicamenteux.¹⁹ Ces défauts peuvent être facilement maîtrisés dans les essais eux-mêmes (au travers de modifications dans la conduite de l'étude si nécessaire) et dans leurs comptes-rendus. Cependant, certains problèmes spécifiques, tels que le choix de l'intervention témoin, l'absence de possibilité d'effectuer l'étude en double-aveugle, l'influence du soignant et les problèmes de standardisation peuvent être difficiles à résoudre lorsque l'on évalue un traitement non médicamenteux.

Intervention témoin

Lorsque l'on évalue un traitement non médicamenteux, il peut être difficile, voire

impossible, de constituer un groupe avec placebo ou intervention factice, pour des raisons éthiques ou techniques. Les considérations éthiques sont d'une grande importance dans les essais évaluant les procédures chirurgicales quand les interventions factices sont prohibitives.^{6,28} Dans les essais examinés ici, les interventions chirurgicales ont toujours été comparées à d'autres procédures chirurgicales, jamais à un placebo. Récemment, Moseley et coll.,²⁹ évaluant des arthroscopies pour gonarthrose, ont mis en avant les difficultés d'effectuer un essai contre placebo dans le cas de la chirurgie. Dans cette étude, le groupe témoin subissait une intervention d'arthroscopie simulée, dans laquelle une petite incision était pratiquée sans qu'aucun instrument ne soit introduit. Les questions pratiques sont également importantes à considérer ; exécuter une intervention factice nécessite souvent une certaine ingéniosité. Par exemple, pour effectuer des essais en double-aveugle contrôlés contre placebo évaluant l'acupuncture, Streitberger et Kleinhenz³⁰ ont mis au point une aiguille-placebo d'acupuncture donnant aux patients une sensation de pénétration de la peau sans que celle-ci ait réellement lieu. Il apparaît donc important de développer des recherches dans ce domaine.

Ignorance du traitement attribué

Empêcher les patients et les soignants de connaître le traitement attribué est généralement possible dans les essais médicamenteux, grâce à l'utilisation d'un placebo ayant la même apparence que le traitement actif, mais cela est souvent impossible dans les essais évaluant des traitements non médicamenteux. Bien que certains investigateurs aient proposé des solutions, comme d'utiliser des pansements pour plaie chirurgicale standardisés pour évaluer la cholécystectomie sous cœlioscopie,³¹ les chirurgiens savent habituellement quelle intervention ils ont pratiqué et les patients savent habituellement quel programme de rééducation ils subissent.

Pour éviter ces biais, des efforts peuvent être faits pour s'assurer que les personnes chargées de l'évaluation restent dans l'ignorance du traitement attribué. L'utilisation du plan expérimental d'étude PROBE (*Prospective Randomized Open Blinded End-point*), basé sur des critères de jugements évalués en aveugle, pourrait être une alternative.³² Cependant, les traitements de la coxarthrose et de la gonarthrose sont habituellement évalués à partir des symptômes rapportés par les patients, et si les patients connaissent leur traitement, les personnes chargées de l'évaluation ne peuvent pas ignorer celui-ci. Les résultats de plusieurs études n'ayant pas été effectuées en double-aveugle ont abouti à des estimations exagérées des effets du traitement,^{3,33} tandis que les résultats d'autres études ne montraient pas d'effet.²⁷ Ces discordances peuvent peut-être être expliquées par

Tableau 2. Caractéristiques d'essais sélectionnés évaluant des traitements médicamenteux et des traitements non médicamenteux*.

Caractéristiques	Essais évaluant des traitements non médicamenteux		Essais évaluant des traitements médicamenteux	
	Chirurgie (n = 17)	Rééducation (n = 23)	Médicaments oraux (n = 41)	Injections intra-articulaires (n = 14)
Randomisation				
Méthode utilisée pour obtenir la séquence de randomisation				
Correcte	5 (29,4)	15 (65,2)	19 (46,3)	3 (21,4)
Incorrecte	1 (5,9)	1 (4,4)	0	0
Non indiquée	11 (64,7)	7 (30,4)	22 (53,7)	11 (78,6)
Dissimulation de la répartition des traitements				
Correcte	0	6 (26,1)	10 (24,4)	1 (7,1)
Incorrecte	8 (47,1)	5 (21,7)	1 (2,4)	0
Non indiquée	9 (52,9)	12 (52,2)	30 (73,2)	13 (92,9)
Reproductibilité de l'intervention	10 (58,8)	22 (95,6)	41 (100)	13 (92,9)
Individualisation	3 (17,6)	11 (47,8)	2 (4,9)	0 (0,0)
Influence des compétences du soignant	17 (100)	16 (69,6)	0 (0,0)	14 (100)
Intervention dans le groupe témoin				
Placebo	0	10 (43,5)	28 (68,3)	11 (78,6)
Traitement actif chez les témoins	17 (100)	9 (39,1)	13 (31,7)	2 (14,3)
Traitement habituel ou groupe témoin sur liste d'attente	0	4 (17,4)	0	1 (7,1)
Effet placebo potentiellement similaire	17 (100)	14 (60,9)	41 (100)	13 (92,9)
Ignorance du traitement attribué				
Patients	4 (23,5)	6 (26,1)	41 (100)	12 (85,7)
Soignants	0	2 (8,7)	41 (100)	3 (21,4)
Personnes évaluant les résultats	7 (41,2)	18 (78,3)	41 (100)	13 (92,9)
Analyse				
AIT effectuée en totalité†	3 (17,6)	6 (26,1)	15 (36,6)	5 (35,7)
AIT modifiée‡	1 (5,9)	4 (17,4)	14 (34,1)	2 (14,3)
Analyse per-protocole effectuée	11 (64,7)	12 (52,2)	10 (24,4)	6 (42,9)
Peu clair	2 (11,8)	1 (4,3)	2 (4,9)	1 (7,1)
Indication du calcul du nombre de sujets à inclure ou de la puissance de l'étude	1 (5,9)	10 (43,5)	29 (70,7)	10 (71,4)

Abbreviations : AIT = analyse en intention de traiter.

* Toutes les données sont exprimées en nombre (pourcentage).

† Tous les participants randomisés ont été inclus dans l'analyse selon le groupe qui leur a été attribué à l'origine par le tirage au sort.

‡ Les analyses ont exclu les participants n'ayant jamais pris leur traitement ou n'ayant jamais été évalués alors qu'ils prenaient leur traitement.

l'hétérogénéité des personnes ignorant le traitement attribué et par les critères de jugement évalués. L'ignorance du traitement est particulièrement importante à considérer quand les critères de jugement impliquent des symptômes rapportés par les patients, tels qu'une douleur, mais elle est moins conséquente pour des critères objectifs, tel que le décès, du fait de faibles biais de détection.

Standardisation et influence du soignant

Les traitements non médicamenteux sont habituellement complexes, ce qui veut dire qu'ils comprennent plusieurs composantes et/ou qu'ils impliquent plusieurs professionnels de santé, et ils sont habituellement individualisés. La composante active de telles interventions est donc difficile à identifier et les traitements sont difficilement reproductibles. Une standardisation détaillée de l'intervention est nécessaire et la qualité technique de l'intervention devrait être évaluée.³⁴ Contrairement aux traitements médicamenteux dans lesquels l'influence du professionnel de santé peut généralement être considérée comme secondaire, dans les traitements non médicamenteux, le soignant fait partie intégrante du traitement. Le succès de l'intervention dépend de ses compétences, de son expérience et de sa formation. Les variations de compétences des soignants dans chaque groupe de traitement peuvent être confondues avec l'effet du traitement.³⁵ La plupart des traitements non médicamenteux, spécialement les interventions chirurgicales, impliquent des procédures complexes et la qualité de la performance nécessite sa fréquente répétition.³⁶ Aucun essai inclus dans notre étude n'a évalué la courbe d'apprentissage et seulement 6 articles se sont intéressés à l'expérience du soignant.

Chirurgie

Parmi les articles portant sur les traitements non médicamenteux, ce sont ceux qui traitaient de chirurgie qui ont eu les plus bas scores de qualité. Des défauts méthodologiques, tels que des comptes-rendus insuffisants concernant les éléments du graphique CONSORT, la dissimulation correcte de la répartition des traitements et la description détaillée de l'intervention, ne devraient pas être plus difficiles à résoudre que dans les essais évaluant d'autres traitements non médicamenteux. Cependant, les essais évaluant les interventions chirurgicales présentent aussi des problèmes spécifiques. Dans ces essais, les soignants peuvent toujours influencer l'effet du traitement et ne sont jamais dans l'ignorance du traitement attribué. Dans les articles analysés, les interventions chirurgicales n'ont jamais été comparées à un placebo, probablement pour des raisons éthiques. Par exemple, les patients inclus dans les essais évaluant les interventions chirurgicales, qui sont irréversibles, n'auront pas l'occasion de bénéficier des résultats de l'essai.

Limites de cette étude

Cette étude était limitée par le fait que nous examinons seulement les comptes-rendus des ECR et non les essais eux-mêmes. L'absence de compte-rendu des méthodes d'un essai ne signifie pas nécessairement que les investigateurs n'ont pas effectué celles-ci. Des défauts méthodologiques peuvent se situer dans le rapport de l'essai plutôt que dans sa conduite.³⁷

Cette étude nous permet de mettre en avant les difficultés de l'évaluation des traitements non médicamenteux. Des améliorations pourraient être apportées en utilisant une méthode correcte pour obtenir la séquence de randomisation et pour dissimuler la répartition des traitements, et en pratiquant une analyse en intention de traiter. Cependant d'autres problèmes méthodologiques sont plus difficiles à résoudre : le choix de l'intervention témoin ; l'ignorance du traitement pour les patients, les soignants et les personnes chargées de l'évaluation ; la standardisation des interventions ; et la prise en compte des compétences du soignant. Enfin, les outils utilisés pour mesurer la qualité, tels que l'échelle de Jadad ou la liste de Delphi, ne sont probablement pas appropriés pour évaluer la qualité des articles sur les traitements non médicamenteux car ils sont centrés sur la randomisation, l'ignorance du traitement attribué et l'analyse en intention de traiter et ne prennent pas en compte des problèmes méthodologiques additionnels, tels que les courbes d'apprentissage, la reproductibilité et l'évaluation de la qualité de l'intervention. Etant donné que les traitements non médicamenteux représentent une gamme large de thérapeutiques disponibles pour les patients, il est important d'améliorer la recherche dans ce domaine et de développer des outils spécifiques appropriés, mais non moins rigoureux, pour évaluer les essais portant sur ce type de traitements. Malgré les défis lancés par les études évaluant les traitements non médicamenteux, ces dernières doivent répondre aux mêmes exigences de qualité que celles qui sont appliquées aux essais évaluant les traitements médicamenteux.

Contributions des auteurs : *conception et schéma de l'étude :* Boutron, Tubach, Giraudeau, Ravaud.

Recueil des données : Boutron, Tubach.

Analyse et interprétation des données : Boutron, Tubach, Giraudeau, Ravaud.

Rédaction du manuscrit : Boutron.

Revue critique du manuscrit : Tubach, Giraudeau, Ravaud.

Expertise statistique : Boutron, Tubach, Giraudeau.

Obtention du financement : Boutron.

Supervision de l'étude : Tubach, Giraudeau, Ravaud.

Financement/Soutien : ce travail a été soutenu par une bourse de la Caisse Nationale d'Assurance Maladie des Professions Indépendantes (CANAM), St Denis, France, et l'Assistance Publique des Hôpitaux de Paris.

BIBLIOGRAPHIE

1. Prescott RJ, Counsell CE, Gillespie WJ, et al. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technol Assess.* 1999; 3:1-143.
2. Guyatt GH, Sackett DL, Cook DJ, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature, II: how to use an article about therapy or prevention, A: are the results of the study valid? *JAMA.* 1993;270:2598-2601.

3. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA.* 1995;273:408-412.
4. Clark HD, Wells GA, Huet C, et al. Assessing the quality of randomized trials: reliability of the Jadad scale. *Control Clin Trials.* 1999;20:448-452.
5. Badley EM. The impact of disabling arthritis. *Arthritis Care Res.* 1995;8:221-228.
6. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. *BMJ.* 2002;324:1448-1451.
7. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ.* 1996; 312:1215-1218.
8. Ravaud P, Moulinier L, Giraudeau B, et al. Effects of joint lavage and steroid injection in patients with osteoarthritis of the knee: results of a multicenter, randomized, controlled trial. *Arthritis Rheum.* 1999;42:475-482.
9. Ike RW, Arnold WJ, Rothschild EW, Shaw HL. Tidal irrigation versus conservative medical management in patients with osteoarthritis of the knee: a prospective randomized study. *J Rheumatol.* 1992;19:772-779.
10. Bradley JD, Heilman DK, Katz BP, Gsell P, Wallick JE, Brandt KD. Tidal irrigation as treatment for knee osteoarthritis: a sham-controlled, randomized, doubleblinded evaluation. *Arthritis Rheum.* 2002;46:100-108.
11. Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. *J Arthroplasty.* 1995;10:133-140.
12. Kreder HJ, Deyo RA, Koepsell T, Swiontkowski MF, Kreuter W. Relationship between the volume of total hip replacements performed by providers and the rates of postoperative complications in the state of Washington. *J Bone Joint Surg Am.* 1997;79:485-494.
13. Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *JAMA.* 2002;287:2805-2808.
14. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996;17:1-12.
15. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol.* 1998;51:1235-1241.
16. Guyatt G, Rennie D. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice.* Chicago, Ill: AMA Press; July 2001:58.
17. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134:663-694.
18. Schulz KF, Grimes DA. Blinding in randomized trials: hiding who got what. *Lancet.* 2002;359:696-700.
19. Hill CL, LaValley MP, Felson DT. Secular changes in the quality of published randomized clinical trials in rheumatology. *Arthritis Rheum.* 2002;46:779-784.
20. van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA.* 1998;280:234-237.
21. Justice AC, Berlin JA, Fletcher SW, Fletcher RH, Goodman SN. Do readers and peer reviewers agree on manuscript quality? *JAMA.* 1994;272:117-119.
22. Verhagen AP, de Vet HC, Vermeer F, et al. The influence of methodologic quality on the conclusion of a landmark meta-analysis on thrombolytic therapy. *Int J Technol Assess Health Care.* 2002;18:11-23.
23. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. *JAMA.* 1996;276:637-639.
24. Hollis S, Campbell F. What is meant by intention to treat analysis? survey of published randomised controlled trials. *BMJ.* 1999;319:670-674.
25. Huiwiler-Muntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA.* 2002;287:2801-2804.
26. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA.* 2002;287:2973-2982.
27. Moher D, Pham B, Jones A, et al. Does quality of reports of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352:609-613.
28. Ridgway PF, Darzi AW. Placebos and standardising new surgical techniques. *BMJ.* 2002;325:560.
29. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med.* 2002;347:81-88.
30. Streitberger K, Kleinhenz J. Introducing a placebo needle into acupuncture research. *Lancet.* 1998;352:364-365.
31. Majeed AW, Troy G, Nicholl JP, et al. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet.* 1996;347:989-994.
32. Hansson L, Hedner T, Dahlof B. Prospective Randomized Open Blinded End-point (PROBE) study: a novel design for intervention trials. *Blood Press.* 1992;1:113-119.
33. Noseworthy JH, Ebers GC, Vandervoort MK, Farquhar RE, Yévis E, Roberts R. The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology.* 1994;44:16-20.
34. Feldon SE, Scherer RW, Hooper FJ, et al. Surgical quality assurance in the Ischemic Optic Neuropathy Decompression Trial (IONDT). *Control Clin Trials.* 2003;24:245-254.
35. Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomized controlled trials. *Stat Med.* 1999;18:2605-2615.
36. Russell I. Evaluating new surgical procedures. *BMJ.* 1995;311:1243-1244.
37. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol.* 2002;55:783-786.