A Systematic Review of Conflicting Meta-Analyses in Orthopaedic Surgery

Patrick Vavken MD, MSc, Ronald Dorotka MD

Received: 10 August 2008 / Accepted: 13 February 2009 / Published online: 28 February 2009
© The Association of Bone and Joint Surgeons 2009

Abstract Meta-analyses are important evaluations in orthopaedic surgery, not only to create clinical guidelines, but also because their findings are included in public health and health policy decision making. However, with increasing numbers of meta-analyses, discordant and frankly conflicting conclusions have been reported. We searched for conflicting meta-analyses, ie, those arriving at different conclusions despite following the same research question, identified potential reasons for these differences, and assessed the statistical significance and clinical importance of differences. We identified conflicting meta-analyses on graft choice in ACL reconstruction and the use of hyaluronic acid. We found significant differences in individual results only for meta-analyses on hyaluronic acid, but the 95% confidence intervals of the magnitude of differences included values as much as 40% for ACL meta-analyses. However, our findings suggest most conflicts derive from differences in the interpretation of pooled results rather than in the actual results. Thus conclusions and interpretations from meta-analyses should be scrutinized as critically as those from any other type of study and subjected to reassessment if deemed necessary.

Each author certifies that he has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

P. Vavken (✉)
Department of Orthopedic Surgery, Children's Hospital Boston,
Harvard Medical School, 300 Longwood Avenue, Enders 1016,
Boston, MA 02115, USA
e-mail: patrick.vavken@childrens.harvard.edu

P. Vavken, R. Dorotka
Department of Orthopedic Surgery, Medical University of
Vienna, Vienna, Austria

Introduction

Evidence-based medicine, commonly defined as the conscientious, explicit, and judicious use of current best evidence, is gaining popularity in orthopaedic research. Usually evidence derives from clinical studies, whose results are weighted based on the level of evidence, determined by methodologic specifics. The highest level of evidence is attributed to randomized controlled trials performed to appropriate methodologic standards and reported according to the CONSORT statement, and meta-analyses of such trials performed to appropriate methodologic standards and reported according to the QUOROM statement [20, 21]. The findings produced in such studies are incorporated into clinical guidelines, serve for public health decisions, and can lead to changes in health economy and health policy [4, 9]. Good examples of such processes are the decision concerning reimbursement of hyaluronic acid, especially in Europe, and recommendations concerning prostate and breast cancer screening. This reflects the importance of such studies.

The popularity of meta-analyses has grown, especially during the last decade with 302 entries under the MeSH-term "meta-analysis" in PubMed in 1990, 1301 in 2000, 2885 in 2005, and finally 3095 publications in 2008, although these techniques have been available for some time [32]. Recent developments in personal computers and statistical software have made these methods available for practically everyone who has a computer and access to the Internet, and the emergence of evidence-based medicine and evidence-based medicine platforms created further interest in this area of research. However, with an increasing number of meta-analyses, disagreement and frankly conflicting conclusions have occurred among meta-analyses studying the same questions and pooling data from similar backgrounds, thus

raising questions regarding the reliability of this current best evidence [6]. Given the importance of meta-analyses and their findings in clinical medicine, public health, health economics, and health policy, such conflict might have important consequences, considering for example, again, reimbursement for hyaluronic acid treatment, which has been discouraged because of negative results in some meta-analyses [1, 18] despite the fact that others [3, 35] concluded there were beneficial effects, as did randomized controlled trials for certain preparations [2, 29, 30].

We therefore posed four objectives: (1) to identify conflicting meta-analyses in orthopaedic research; (2) to ascertain whether differences in study quality or study characteristics explain such conflicts; (3) to assess the statistical significance and clinical importance, ie, the size, of the actual differences in results in meta-analyses with conflicting conclusions; and (4) to use our findings to identify among conflicting meta-analyses those that provide the most relevant answers to their question.

Search Strategies and Criteria

To identify conflicting meta-analyses, we performed an online systematic literature review of orthopaedic meta-analyses using variations of the terms “orthopedic,” “orthopedics,” and “orthopedic procedures” and “meta-analysis” in PubMed and the Cochrane Library (Fig. 1). All searches were concluded by February 1, 2008. We defined meta-analyses eligible for inclusion in our study as those performing a systematic review and formal data synthesis of relevant data on a defined study question. Studies of orthopaedic treatments were eligible, whereas studies on medical, radiologic, or anesthesiologic procedures in orthopaedic patients were excluded. In addition to the database search, we interviewed four orthopaedic surgeons involved in evidence-based medicine, epidemiology, or biostatistics to identify further conflicting meta-analyses. We identified 218 eligible published meta-analyses and 61 Cochrane reviews. All meta-analyses were categorized according to topic and study questions. Conflicting meta-analyses were defined as those investigating the same study question but arriving at a different conclusion or treatment recommendation (Fig. 1).

To identify conflict arising from differences in study quality, we (PV, RD) used the Oxman-Guyatt [20, 21] and the QUOROM [15] indices to assess the overall methodologic quality of meta-analyses. The Oxman-Guyatt index is an instrument to measure the methodologic quality of review articles ranging from 1 (extensively flawed) to 7 (minimally flawed) points [20, 21], whereas the QUOROM index derives from the QUOROM statement and is a measure of the quality of reporting of meta-analyses

ranging from 0 (worst) to 18 (best) points [15]. Additionally to those quantitative methods, we defined five questions to qualitatively assess potential sources of conflict arising from differences in materials and methods in meta-analyses. First, we asked whether the rationale of repeating the meta-analysis was reported including reference to prior meta-analyses. Although this is not a formal parameter of study quality, a thorough systematic review of the literature and inclusion of all findings is the bedrock of a meticulous meta-analysis. Thus detection and discussion of earlier meta-analyses can be considered a surrogate for diligent handling. Second, we asked whether the same studies and/or studies of the same level of evidence were included. Third, we asked whether the same end points were studied. Fourth, we asked whether the same methods were used in the analytical approach. Here we looked specifically at how publication bias was ruled out, how study quality was assessed, how heterogeneity was tested and accounted for, and how data were pooled. Our final question was whether there were other apparent problems.

After assessing potential sources of conflict in conclusions, we asked whether these conclusions were supported by the results of meta-analyses. To assess the significance and the magnitude of differences in presented end points between conflicting meta-analyses, Cochran's chi square test for heterogeneity, or Q test for short, and I^2 index were used. The Q test is a validated instrument to test the significance of differences between study results in meta-analyses. An important shortcoming of this test is that its power is low in small samples [13, 14]. Thus we set the threshold of significance for the Q test at 0.1 instead of 0.05, to adjust for the small size of our sample and to avoid missing differences because of false negative results. The Q test assumes statistical independence, however, because the studied meta-analyses have at least some overlap in their primary studies, this assumption might be violated. Such a violation leads to too small p values. In addition to the significance of differences, the I^2 index gives an estimate of the actual magnitude of between meta-analysis variability relative to the total amount of variability, which also includes the play of chance [14]. The I^2 index also depends on sample size; thus, we calculated the 95% confidence intervals for the I^2 index based on the method by Higgins et al. [14, 15] to account for this potential problem. This method is robust to low sample sizes in contrast to the Q test and the I^2 index. Q tests, I^2 indices, and confidence intervals were calculated using Intercooled Stata[®] 10 (StataCorp LP, College Station, TX).

To identify among meta-analyses, those that provide the most relevant answers to their question, we used the following algorithm. First, meta-analyses were ranked by study quality as given by the Oxman-Guyatt and the QUOROM indices. If there were conflicts among meta-

Fig. 1 The flow of studies in the systematic search of conflicting meta-analyses is shown. Studies were excluded if they were not about orthopaedic procedures, eg, anesthesiologic studies in orthopaedic patients. CL = Cochrane Library.

analyses of high quality, we used the five questions of our qualitative assessment as described above to identify and address potential sources of such conflict. If conflicting conclusions remained, we used the significance and magnitude of the differences in end points to test whether conflicts are supported.

Results

We found conflicting meta-analyses in two topics: choice of graft in ACL reconstruction ($n = 7$) (Table 1) [5, 8, 10, 12, 28, 33, 36], and hyaluronic acid for treating osteoarthritis of the knee ($n = 5$) (Table 2) [1, 7, 18, 19, 35]. For these two topics, we could extract data on pain for hyaluronic acid, and the Lachman test, pivot shift, and extension lag for graft choice in ACL reconstruction.

It is unlikely study quality was a reason for differences among these studies. The assessment of the quality of the meta-analyses as a potential source or explanation for differing results showed considerable differences in methodologic quality and quality of reporting only for meta-analyses of graft choice for ACL reconstruction (Table 3), but fairly high and evenly distributed values for meta-analyses of hyaluronic acid. Our first question in the qualitative assessment of sources of conflicts among meta-analyses of ACL graft choice showed only two studies reported the rationale of repeating the meta-analysis. Furthermore, the levels of evidence among the included primary studies in these meta-analyses were rather low and vary among studies, especially with time. Also there was variation in the numbers of studied end points, but all studies included stability (Table 1). Methodologically, most meta-analyses were rather low quality. Assessment of

Table 1. Characteristics of meta-analyses on graft choice in ACL reconstruction

Study	Journal	Year	Included primary studies	Included patients	End points (as reported)	Interpretation	Direction
Biau et al. [5]	British Medical Journal	2006	19 RCTs (from 18 cohorts) from MEDLINE, WebSPIRS, Science Citation Index, Current Contents, CCTR until March 14, 2005	1512 patients (765 BTB, 747 HT)	Lachman, pivot shift, anterior knee pain, kneeling test, extension loss, graft failure, arthrometer	Less donor site morbidity for HT, minimal evidence for better stability after BTB	Favors HT
Forster and Forster [8]	Knee	2005	6 prospective, randomized or quasi-randomized trials from MEDLINE, EMBASE, CCTR	475 patients (235 BTB, 240 HT)	Extension loss, flexion loss, Lachman, pivot shift, arthrometer, return to sports, clinical knee scores, isokinetic muscle testing, patellofemoral pain	Extension lag and anterior knee pain significantly higher with BTB, less quadriceps strength and worse pivot shift with HT, no difference in Lachman, return to sports, clinical scores and rate of complications	Neutral
Goldblatt et al. [12]	Arthroscopy	2005	11 RCTs from MEDLINE	1039 patients (515 BTB, 524 HT)	Anterior knee pain, activity-related pain, giving-way, flexion loss, extension loss, Lachman, pivot-shift, patellofemoral crepitus, IKDC score, KT-1000 arthrometer, return to preinjury level of activity, flexion strength, extension strength, swelling, graft rupture, Lysholm score, Tegner score, meniscus surgery, complications	Better results in Lachman, pivot shift, KT-1000, and flexion with BTB	Favors BTB
Prodromos et al. [28]	Arthroscopy	2005	32 BTB and 32 HT series (from 35 studies) from PubMed	1594 HT, 1808 BTB (patients with data on stability)	Stability (within 2 mm of contralateral knee in KT-1000 arthrometer assessment)	Better stability with HT, dependent on fixation (EndoButton® is most reliable, better than screw)	Favors HT

Table 1. continued

Study	Journal	Year	Included primary studies	Included patients	End points (as reported)	Interpretation	Direction
Thompson et al. [33]	American Journal of Orthopedics	2005	2 RCTs from PRE-MEDLINE and MEDLINE until May 2001	157 BTB, 151 HT	Overall knee condition, IKDC, patient satisfaction, activity level, motion loss, patellofemoral pain, harvest site symptoms, thigh atrophy, kneeling pain, Lysholm score, Lachman, pivot shift	High risk of extension lag with BTB, higher risk of abnormal IKDC scores, no significant functional differences	Favors HT, function neutral
Freedman et al. [10]	American Journal of Sports Medicine	2003	21 BTB studies, 13 HT studies, all from MEDLINE	1348 BTB, 628 HT	Graft failure, KT-1000 arthrometer, side-to-side difference in laxity, pivot shift, loss of motion, patient satisfaction, return to previous level of sports, complications, knee scores	Significantly less graft failures and better stability with BTB, but increased risk of anterior knee pain	Favors BTB, but anterior knee pain
Yunes et al. [36]	Arthroscopy	2001	4 RCTs from MEDLINE until 1997	229 BTB, 183 HT	Performance level, KT arthrometer, Lachman, pivot shift, range of motion, complications, failures	Higher postoperative activity and better stability with BTB	Favors BTB

RCT = randomized, controlled trial; CCTR = Cochrane Controlled Trial Register; BTB = bone-tendon-bone; HT = hamstring tendon; IKDC = International Knee Documentation Committee.

Table 2. Characteristics of meta-analyses on hyaluronic acid for management of osteoarthritis of the knee

Study	Journal	Year	Included primary studies	Included patients	End points (pooled as)	Interpretation	Direction
Modawal et al. [19]	The Journal of Family Practice	2005	9 RCTs from MEDLINE, CCTR, Agency for Healthcare Research and Quality	1207 patients (593 treated, 614 controls)	Knee pain at rest and on activity	Moderate effect from 5–7 and from 8–10 weeks, no effect after 15–22 weeks	Favors HA
Arrich et al. [1]	Canadian Medical Association Journal	2005	22 RCTs from MEDLINE, EMBASE, CINAHL, BIOSIS	3222 patients (1635 treated, 1587 controls)	Pain at rest, pain during exercise, joint function, adverse events	No improvement in pain, no change in function, 1%–15% increase in rate of complications	Against HA
Wang et al. [35]	Journal of Bone and Joint Surgery American Volume	2004	20 RCTs from MEDLINE, EMBASE, Cochrane Library	1647 knees (818 treated, 829 controls)	Pain with activity, pain without activity, function (all transformed)	Significantly improved pain and function, no difference in complications	Favors HA
Espallargues and Pons [7]	International Journal of Health Technology Assessment	2003	14 studies (7 RCTs, 6 cohorts, 1 cross-sectional study) from MEDLINE, EMBASE, HealthSTAR, Current Concepts, Cochrane Library	1735 patients	Pain, delay of prosthesis, adverse effects	Short-term improvement in pain and function, long-term effect or delay of prosthesis not proven	Neutral to favors HA
Lo et al. [18]	Journal of the American Medical Association	2003	22 RCTs from MEDLINE, Cochrane Library	2949 patients	Effect size of pain reduction	Low effect (0.19), clinical meaning questionable	Neutral to against HA

RCT = randomized, controlled trial; CCTR = Cochrane Controlled Trial Register; HA = hyaluronic acid.

Table 3. Potential reasons for conflicts in meta-analyses on graft choice in ACL reconstruction

Study	Journal	Year	QUOROM Score	Oxman-Guyatt Index	Rationale for repeating the analysis*	Earlier studies quoted	Heterogeneity, sensitivity analyses	Other potential problems
Biau et al. [5]	British Medical Journal	2006	18	7	"... provide up to date knowledge ..."	Forster and Forster [8]	Yes	
Forster and Forster [8]	Knee	2005	13	3	"... the results of 4HS were not analyzed separately ..."	Yunes et al. [36] Yunes et al. [36] Freedman et al. [10]	Yes	
Goldblatt et al. [12]	Arthroscopy	2005	16	2	"... surgical techniques, rehabilitation protocols, and methods of fixation have evolved. Additional studies are available ..."	Yunes et al. [36] Freedman et al. [10]	Yes	
Prodromos et al. [28]	Arthroscopy	2005	11	1	Focus on EndoButton® and 4-string technique	Yunes et al. [36] Freedman et al. [10]	No	Simpson's paradox [†]
Thompson et al. [33]	American Journal of Orthopedics	2005	9	2	Not reported	Yunes et al. [36]	No	
Freedman et al. [10]	American Journal of Sports Medicine	2003	14	2	Not reported	Yunes et al. [36]	Yes	
Yunes et al. [36]	Arthroscopy	2001	13	3	†	†	Yes	

* As stated in text; †Simpson's paradox = ranking of proportions in individual groups is reversed in the combined total after mere addition; ‡ first meta-analysis.

Table 4. Methods used in the included meta-analyses to assess and pool primary studies

Topic	Study	Year	Publication bias		Study quality by		Heterogeneity			Pooling	
			Funnel plot	Egger's linear regression	Score	Components	Galbraith plot	Q test	I ² index	Sensitivity analysis	Fixed effects
ACL graft choice	Biau et al. [5]	2006			x			x			x
	Forster and Forster [8]	2005						x		x	
	Goldblatt et al. [12]	2005							x		x
	Prodromos et al. [28]	2005							x		
	Thompson et al. [33]	2005									
	Freedman et al. [10]	2003									
Hyaluronic acid	Yunes et al. [36]	2001						x		x	
	Modawal et al. [19]	2005		x	x		x		x		x
	Arrich et al. [1]	2005		x		x		x	x		x
	Wang et al. [35]	2004	x		x				x		x
	Espallargues and Pons [7]	2003			x					x	x
	Lo et al. [18]	2003	x	x				x			x

publication bias was neglected completely, and only Biau et al. [5] tested study quality. Also, heterogeneity was rarely adequately considered (Table 4). Finally, the validity of the methods used to pool data occasionally was unclear. Prodromos et al. [28] and Thompson et al. [33], for example, merely added numbers from the primary studies to obtain overall results. This, however, can lead to biased results owing to Simpson's paradox (ie, the outcomes appear reversed when the groups are combined), particularly if the treatment allocation is not equal in all studies. Also, there seemed to be important differences in the surgical procedures, such as the type of fixation used. The qualitative assessment of meta-analyses on hyaluronic acid in osteoarthritis showed two did and two did not report a rationale for repeating the analysis (Table 5). Although Wang et al. [35], Arrich et al. [1], and Lo et al. [18] included essentially the same primary studies, Modawal et al. [19] included fewer studies, considering only a selection of hyaluronic acid preparations. All authors but Espallargues and Pons [7] included only randomized controlled trials and performed sensitivity analyses to include heterogeneity in their analyses. The same end points were studied but analyzed in different ways. Wang et al. [35] standardized his pooled estimates on maximum scale intensity and trial duration, thus making direct comparison with the pooled estimates in the other studies difficult. All studies considered publication bias, study quality, and heterogeneity in their analyses (Table 4).

Statistical inference suggested no significant differences in the numerical results of conflicting meta-analyses other than for hyaluronic acid (Table 6). Quantitative assessment of the magnitude of difference among the results of conflicting meta-analyses by the I² index corroborated these findings. However, the 95% confidence intervals, adjusted for small sample size, suggested there might be as much as 41% of true difference beyond random variation in results for meta-analyses of ACL grafting (Table 4).

Our last objective was to use our findings to identify among conflicting meta-analyses those that provided the most relevant answers to their question. For ACL graft choice the studies by Biau et al. [5], Forster and Forster [8], and Goldblatt et al. [12] had the highest methodologic quality, but still presented conflicting conclusions (Tables 1 and 3). As mentioned above, the qualitative assessment of studies showed differences only in the surgical procedures, suggesting confounding by this variable [34]. Another study, which is not a primary meta-analysis and thus was not included in this study, found that after adjustment for type of fixation, hamstring tendon produces better postoperative outcome in pain, and at least equal outcomes in stability [27]. The existing data suggested the most relevant answer is that a hamstring tendon graft produces a better outcome, to a small extent, than a

Table 5. Potential reasons for conflicts in meta-analyses on hyaluronic acid

Study	Journal	Year	QUOROM Score	Oxman-Guyatt Index	Rationale for repeating the analysis*	Earlier studies quoted	Heterogeneity, sensitivity analyses	Other potential problems
Modawal et al. [19]	The Journal of Family Practice	2005	15	5	"... stringent test of the efficacy of viscosupplementation for relieving knee pain from osteoarthritis with a meta-analysis that includes only data from randomized, double-blinded, controlled trials of hyaluronic acid that measured pain using a visual analogue scale (VAS), the most widely accepted method for pain evaluation."	Lo et al. [18]	Yes	
Arrich et al. [1]	Canadian Medical Association Journal	2005	18	7	"used different approach to data synthesis and interpretation ... allocated trial data ... to 3 outcome groups ... assumed would be relevant for patients with osteoarthritis"	Wang et al. [35] Lo et al. [18]	Yes	
Wang et al. [35]	Journal of Bone and Joint Surgery American Volume	2004	17	6	Not reported	None	Yes	"efficacy scores" = effect standardized on trial duration and scale used, comparison with other meta-analyses is difficult
Espallargues and Pons [7]	International Journal of Health Technology Assessment	2003	16	5	Not reported	None	No	
Lo et al. [18]	Journal of the American Medical Association	2003	16	6	†	†	Yes	

* As stated in text; † first meta-analysis.

Table 6. Significance and amount of difference in end points among conflicting meta-analyses

Topic	End point	Measure	Q test (p value)*	I ² index	95% CI (I ²) [†]
ACL reconstruction	Lachman test	Relative risk [‡]	0.952	0%	0%–25.4%
	Pivot shift	Relative risk [‡]	0.94	0%	0%–28.3%
	Extension lag	Relative risk [‡]	0.915	0%	0%–41.4%
Hyaluronic acid	Pain	Pain scale	< 0.01	66%	38.6%–78.6%

* Because of the low power of the Q test in small samples, $p < 0.1$ is considered significant; [†]according to Higgins and Thompson [14]; [‡]relative risk: bone-tendon-bone as baseline; the lack of independence of pooled results between meta-analyses with large overlap in primary studies might have resulted in too small p values and too narrow 95% CI.

bone-tendon-bone graft. As for hyaluronic acid, study quality was consistently high in all meta-analyses by Oxman-Guyatt and QUOROM indices and by our qualitative assessment. When looking for apparent problems (the last question in our qualitative assessment), we saw that these studies combined distinctly different preparations of hyaluronic acid, which is consistent with clinical heterogeneity and begs the question whether pooling was justified. The statistically significant difference between the presented end points in the meta-analyses suggests that there is a true difference. The most relevant answer is that individual preparations of hyaluronic acid should be considered one at a time or in subgroups.

Discussion

Meta-analyses have gained meaning in orthopaedic surgery as an instrument for evidence-based clinical decision making, and for public health decisions, by providing high-quality evidence pooled from the current literature. Errors and uncertainties or even conflicting conclusions in meta-analyses, in turn, will cause confusion and might even precipitate severely problematic consequences. This potential problem led to this study, which focused on conflicting Level 1 evidence from orthopaedic meta-analyses. Specifically, we had four objectives. First, we wanted to identify any conflicting meta-analyses in orthopaedic surgery. Our second objective addressed potential sources for conflict. Third, we studied differences in outcomes in conflicting studies to see whether these differences would support conflicting conclusions. Finally, we tried to present the best evidence for each topic, building on our findings.

Our study has some limitations. Most importantly, it builds on what is reported in the literature, which is not necessarily the same as what was done in the included studies. This study is also especially susceptible to publication bias because it is likely many meta-analyses presenting results in conflict with earlier studies were never published. However, we present all studies we were able to identify, and our study can be seen as a comprehensive sample rather than a complete account of all conflicting

meta-analyses without jeopardizing the validity of our findings. Another potential shortcoming derives from our definition of “conflicts.” We only included meta-analyses based on conflict in their conclusions rather than numerical result because we argued the conclusion, as the final product of any study, has more meaning. Also, we were not able to investigate all potential sources of conflict in the scope of this article but focused on the most important ones. We tried to present the most relevant answers, supported by strong evidence, for each topic. Finally, concerning our statistical analysis, there is a lack of independence between pooled estimates because there is overlap in the primary studies from which these pooled estimates were built. Such a lack of independence is a violation of the assumptions of the used Q-test because the data are more similar than the test expects, which in turn could cause spuriously small p values and too narrow confidence intervals.

We were able to identify conflicting meta-analyses, defined as studies arriving at different conclusions despite studying the same literature to answer the same question, in two orthopaedic topics: graft choice in ACL reconstruction and management of osteoarthritis with hyaluronic acid. Poolman et al. [26] also investigated systematic reviews on the latter topic. However, we chose to focus on meta-analyses, which add analytical methods to systematic reviews to arrive at a pooled estimate of treatment effectiveness, instead of merely describing previous results in a systemic fashion. Discordant and openly conflicting meta-analyses have been reported in other fields as well [16]. For example, two meta-analyses of heparin in the prevention of perioperative thrombosis published in the *British Medical Journal* and *Lancet* during the same year led to opposite results [17, 25]. Other reports of conflicting meta-analyses have been mentioned in conjunction with commercial interests; for example, a meta-analysis refuting the effectiveness of selective serotonin reuptake inhibitors in the management of depression after its publication was promptly challenged by a meta-analysis presented by Lilly Industries, coming to a clearly different result and showing high effectiveness for this type of medication [24, 31].

Subsequently, we assessed potential sources of conflict. There were little differences in quality indices for the meta-

analyses on hyaluronic acid, thus precluding this as an explanation for differences. For studies on graft choice, however, considerable differences could be seen, allowing elimination of studies with low scores to come closer to a conclusion. However, both indices we used build on information reported in the papers and thus, as mentioned above, might be biased by differences in the accuracy of reporting as orthopaedic journals might emphasize statistical/methodologic details differently from others or even reduce the level of detail in reporting owing to space limitations. Thus, we added a qualitative part to this assessment, using four questions, chosen from commonly known problems and limitations in primary and repeated meta-analyses. The first and maybe most important question concerned the rationale for repeating the analysis. Although this subject has hardly been studied, repeating or updating rarely (9%) leads to changes in the pooled results of meta-analyses [11, 22, 23]. However, conclusions from meta-analyses have been refuted by subsequently published analyses and even by large randomized controlled trials [6]. Therefore, repeating of meta-analyses might provide new and important changes and should not be discouraged *a priori*. Of further interest was how authors defended their conclusions in relation to prior studies, especially when taking into account that a conflicting, and thus new, conclusion might increase the chances for successful submission for publication. Also, the availability of new evidence might be important in altering conclusions when repeating meta-analyses, however, all the presented meta-analyses were published in a maximum time frame of 5 years. Thus, a more likely source of new evidence is non-English studies, which have been excluded owing to language bias, or publication bias rather than newly published evidence. Our other questions focused mostly on methodologic issues as investigators may choose from a range of methods of how to identify, weigh, and pool data in meta-analyses, which, despite being equally valid, might produce quite different values for pooled results. Studying the methodologies of the included meta-analyses, we saw large between-topic differences but fairly consistent quality in a topic. It is noteworthy that none of the meta-analyses on ACL grafts reported on assessment of publication bias. Such testing, usually accomplished graphically using funnel plots or mathematically using Egger's regression or Begg's ranked correlation, shows the homogeneity of the literature and thus can be seen as the initial step toward assessing the appropriateness of pooling. Also, only one of these meta-analyses assessed study quality. It is important to consider internal and external validity when choosing primary studies for inclusion in a meta-analysis. Internal validity, which is the validity of the used methods to answer the study question, is best analyzed by assessment of the study design. External validity, which is the

generalizability of results or the applicability in a specific population of interest, is best assured by stringent inclusion and exclusion criteria. Stringent inclusion and exclusion criteria in primary studies cannot replace assessment of internal validity. However, the inclusion in meta-analyses of outcomes of assessments of study quality is difficult. Very much like publication bias and study quality, the assessment of heterogeneity was gravely neglected in many meta-analyses of ACL graft choice. Accounting for heterogeneity is of special importance in surgical studies because, compared with pharmaceutical treatments, surgical treatment cannot be completely standardized, and outcomes will vary even if the same surgeon performs the same procedure. Meta-analyses on hyaluronic acid addressed publication bias, study quality, and heterogeneity adequately and according to current standards. Another problem that stands out in all topics, however, is appropriateness of pooling. Especially for hyaluronic acid, the studies in conflict with each other pooled data on different combinations of preparations, suggesting the effectiveness of some is diluted in the lack of effect of others. This is confirmed by findings presented in the systematic review by Bellamy et al. [3].

Finally, after studying potential sources of conflict, we wanted to study the actual differences in numerical outcomes to test whether they would support conflicting conclusions. We used the two parameters quality, or statistical significance, and quantity, or the actual magnitude of difference, independently because a statistically significant difference between outcomes is not necessarily clinically important and a nonsignificant difference might have a clinically meaningful size nevertheless. Especially in meta-analyses, with often quite large sample sizes, standard errors might be very small, thus creating very small *p* values leading to further overemphasis of statistical significance at the expense of clinical importance. Our assessment showed virtually all the numerical results in conflicting meta-analyses were not significantly different from each other. Also, when the magnitudes of differences were assessed using the I^2 index, which is a measure of heterogeneity among results beyond difference attributable to normal, random variation, we found mostly no differences. However, we also calculated 95% confidence intervals for I^2 adjusted for small sample size and could show these intervals allowed for some considerable differences. Our findings suggest much of the actual conflict in conclusions originated from the interpretation of results rather than from their actual values.

The question of which conclusion to trust among conflicting meta-analyses remains to be answered. Jadad et al. [16] suggested an algorithm in 1997 to deal with discordant systematic reviews, building on parameters of study design such as study question, inclusion criteria, and methods of

pooling. We modified this algorithm using the findings from our study and used it to find the best evidence for both studied topics. For hyaluronic acid, all studies are of equal quality and build on equal materials and methods. For this topic, as mentioned above, it seems pooling was not always appropriate. Lo et al. [18] and Arrich et al. [1] found differences in effectiveness attributable to molecular weight in their analysis, although Arrich et al. [1] did not find a significant association. Furthermore, there is strong evidence from randomized controlled trials and meta-analyses for the effectiveness of some hyaluronic acid preparations in the treatment of osteoarthritis [2, 7, 29, 30]. We did not find a best answer, but rather a best question. A similar situation can be seen in ACL grafts. Poolman et al. [27] used the data published by Biau et al. [5] in a cumulative meta-analysis with sensitivity analyses for the type of fixation. Poolman et al. presented clear evidence supporting the use of hamstring grafts with modern fixation technique (EndoButton®) resulting in equal results in stability at lower complication rates than bone-patellar tendon-bone grafts. Also, by using cumulative meta-analysis, this group could show these findings could have been established as early as in 2001. Future studies on hyaluronic acid would benefit from a similar approach.

We have identified meta-analyses with conflicting conclusions in orthopaedic surgery. However, we believe these conflicts derive from interpretation of results rather than from the pooled results, and therefore they might be resolved by assessing whether conclusions are supported by the results and the way those results were generated, which also includes assessment of the appropriateness of pooling. Additionally, although several studies suggest repeating or updating meta-analyses rarely produces changes in results, we found this is not necessarily true. We recommend meta-analyses be treated like any other study and their results be critically scrutinized and reassessed if justified. Repeating meta-analyses in the face of new evidence might produce important new findings and add to our knowledge.

References

1. Arrich J, Piribauer F, Mad P, Schmid D, Klaushofer K, Müllner M. Intra-articular hyaluronic acid for the treatment of osteoarthritis of the knee: systematic review and meta-analysis. *CMAJ*. 2005;172:1039–1043.
2. Bellamy N, Bell MJ, Goldsmith CH, Pericak D, Walker V, Raynauld JP, Torrance GW, Tugwell P, Polisson R. The effectiveness of hylan G-F 20 in patients with knee osteoarthritis: an application of two sets of response criteria developed by the OARSI and one set developed by OMERACT-OARSI. *Osteoarthritis Cartilage*. 2005;13:104–110.
3. Bellamy N, Campbell J, Robinson V, Gee T, Bourne R, Wells G. Viscosupplementation for the treatment of osteoarthritis of the knee. *Cochrane Database Syst Rev*. 2006;CD005321.
4. Bero LA, Jadad AR. How consumers and policymakers can use systematic reviews for decision making. *Ann Intern Med*. 1997;127:37–42.
5. Biau DJ, Tournoux C, Katsahian S, Schranz PJ, Nizard RS. Bone-patellar tendon-bone autografts versus hamstring autografts for reconstruction of anterior cruciate ligament: meta-analysis. *BMJ*. 2006;332:995–1001.
6. Egger M, Dickersin K, Smith GD. Problems and limitations in conducting systematic reviews. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care: Meta-Analysis in Context*. London, UK: BMJ Publishing Group; 2005:43–68.
7. Espallargues M, Pons JM. Efficacy and safety of viscosupplementation with Hylan G-F 20 for the treatment of knee osteoarthritis: a systematic review. *Int J Technol Assess Health Care*. 2003;19:41–56.
8. Forster MC, Forster IW. Patellar tendon or four-strand hamstring? A systematic review of autografts for anterior cruciate ligament reconstruction. *Knee*. 2005;12:225–230.
9. Fox DM. Evidence of evidence-based health policy: the politics of systematic reviews in coverage decisions. *Health Aff (Millwood)*. 2005;24:114–122.
10. Freedman KB, D'Amato MJ, Nedeff DD, Kaz A, Bach BR Jr. Arthroscopic anterior cruciate ligament reconstruction: a meta-analysis comparing patellar tendon and hamstring tendon autografts. *Am J Sports Med*. 2003;31:2–11.
11. French SD, McDonald S, McKenzie JE, Green SE. Investing in updating: how do conclusions change when Cochrane systematic reviews are updated? *BMC Med Res Methodol*. 2005;5:33.
12. Goldblatt JP, Fitzsimmons SE, Balk E, Richmond JC. Reconstruction of the anterior cruciate ligament: meta-analysis of patellar tendon versus hamstring tendon autograft. *Arthroscopy*. 2005;21:791–803.
13. Hardy R, Thompson S. Detecting and describing heterogeneity in meta-analysis. *Stat Med*. 1998;17:841–856.
14. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21:1539–1558.
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560.
16. Jadad AR, Cook DJ, Browman GP. A guide to interpreting discordant systematic reviews. *CMAJ*. 1997;156:1411–1416.
17. Leizorovicz A, Haugh MC, Chapuis FR, Samama MM, Boissel JP. Low molecular weight heparin on the prevention of perioperative thrombosis. *BMJ*. 1992;305:913–920.
18. Lo GH, LaValley M, McAlindon T, Felson DT. Intra-articular hyaluronic acid in treatment of knee osteoarthritis: a meta-analysis. *JAMA*. 2003;290:3115–3121.
19. Modawal A, Ferrer M, Choi HK, Castle JA. Hyaluronic acid injections relieve knee pain. *J Fam Pract*. 2005;54:758–767.
20. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses. *Lancet*. 1999;354:1896–1900.
21. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 2001;357:1191–1194.
22. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. A systematic review identified few methods and strategies describing when and how to update systematic reviews. *J Clin Epidemiol*. 2007;60:1095–1104.
23. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. *Cochrane Database Syst Rev*. 2008;MR000023.
24. Nakielnny J. Effective and acceptable treatment for depression. *BMJ*. 1993;306:1125.
25. Nurmohamed MT, Rosendaal FR, Büller HR, Dekker E, Hommes DW, Vandenbroucke JP, Briët E. Low-molecular weight heparin

- versus standard heparin in general and orthopaedic surgery: a meta-analysis. *Lancet*. 1992;340:152–156.
26. Poolman RW, Abouali JA, Conter HJ, Bhandari M. Overlapping systematic reviews of anterior cruciate ligament reconstruction comparing hamstring autograft with bone-patellar tendon-bone autograft: why are they different? *J Bone Joint Surg Am*. 2007;89:1542–1552.
 27. Poolman RW, Farrokhyar F, Bhandari M. Hamstring tendon autograft better than bone patellar-tendon bone autograft in ACL reconstruction: a cumulative meta-analysis and clinically relevant sensitivity analysis applied to a previously published analysis. *Acta Orthop*. 2007;78:350–354.
 28. Prodromos CC, Joyce BT, Shi K, Keller BL. A meta-analysis of stability after anterior cruciate ligament reconstruction as a function of hamstring versus patellar tendon graft and fixation type. *Arthroscopy*. 2005;21:1202.
 29. Raynauld JP, Goldsmith CH, Bellamy N, Torrance GW, Polisson R, Belovich D, Pericak D, Tugwell P. Effectiveness and safety of repeat courses of hylan G-F 20 in patients with knee osteoarthritis. *Osteoarthritis Cartilage*. 2005;13:111–119.
 30. Raynauld JP, Torrance GW, Band PA, Goldsmith CH, Tugwell P, Walker V, Schultz M, Bellamy N. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. *Osteoarthritis Cartilage*. 2002;10:506–517.
 31. Song F, Freemantle N, Sheldon TA, House A, Watson P, Long A, Mason J. Selective serotonin reuptake inhibitors: meta-analysis of efficacy and acceptability. *BMJ*. 1993;306:683–687.
 32. Thacker SB. Meta-analysis: a quantitative approach to research integration. *JAMA*. 1988;259:1685–1689.
 33. Thompson J, Harris M, Grana WA. Patellofemoral pain and functional outcome after anterior cruciate ligament reconstruction: an analysis of the literature. *Am J Orthop*. 2005;34:396–399.
 34. Vavken P, Culen G, Dorotka R. Management of confounding in controlled orthopaedic trials: a cross-sectional study. *Clin Orthop Relat Res*. 2008;466:985–989.
 35. Wang CT, Lin J, Chang CJ, Lin YT, Hou SM. Therapeutic effects of hyaluronic acid on osteoarthritis of the knee: a meta-analysis of randomized controlled trials. *J Bone Joint Surg Am*. 2004;86:538–545.
 36. Yunes M, Richmond JC, Engels EA, Pinczewski LA. Patellar versus hamstring tendons in anterior cruciate ligament reconstruction: a meta-analysis. *Arthroscopy*. 2001;17:248–257.